RISK - OPTIMIZED SEQUENTIAL SAMPLING PLANS

By
RAHUL SINGH

DEPARTMENT OF INDUSTRIAL AND MANAGEMENT ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

APIRL, 1991

RISK-OPTIMIZED SEQUENTIAL SAMPLING PLANS

A Thesis Submitted

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF TECHNOLOGY

by

Rahul Singh

to the

Department of Industrial and Management Engineering

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

April, 1991

CERTIFICATE

18/4/9

It is certified that the work contained in the thesis entit "*Risk-optimized Sequential Sampling Plans*", by Rahul Singh, has b carried out under my supervision and that this work has not b submitted elsewhere for a degree.

Dr. Tapan P. Bagchi

Department of Industrial and Management Engineering

Indian Institute of Technology, Kanpur

ACKNOWLEDGEMENTS

Working on this thesis has been my first real introduction to the discipline of research. There is nothing like expert guidance in the very beginning to get a novice on the productivity track so that he stays right there throughout his life, and I am glad to say that I was certainly lucky to have run into Dr. Tapan P. Bagchi in the final year of my Bachelors, and had the right mind to stick with him throughout my Masters. The least he has done to me is that he has put me on this productivity track so that I will stay right there throughout my life.

I was certainly lucky to have run into the Department of IME too, which is unique in itself in IIT Kanpur, with nice and helpful colleagues, professors, and staff. One could never get oneself not to work hard because there was always someone who was working harder than you. From Prusty's pipes to Dada's databases to Naik's parametrics to Satya's frequencies to Sanjay's organizations to Sukhram's flexibility to Dangi's CFAs to Paradkar's trains to Modi's echelons to Rao's ghosts, everybody was trying his best and enjoying it. I owe a lot to everybody for making this third semester so involving and interesting, and to our lab, the life support system of our department.

ABSTRACT

The real problem in acceptance sampling is to design or choose sampling procedures that accurately represent the consequences of accepting or rejecting the lot. Most of the acceptance sampling procedures in use today are based on *statistical* (probability) measures, with the decision maker assigning desirable probabilities to the events of accepting the lot at LTPD and of rejecting the lot at AQL. Sequential sampling, a valued approach when the cost of inspection is high, additionally aims at minimizing the expected number of items sampled.

In the present work, optimal sequential procedures for acceptance sampling based on the *risk* consequences of accepting or rejecting a lot are derived and studied. "Risk" here incorporates the true loss of utility (of which monetary loss is a special case) to the decision maker in the *decision-theoretic* sense, so that the plans obtained minimize the total expected risk of accepting/rejecting the lot, making full use of the existing (prior) knowledge of the unknown lot quality. The decision is guided by on one hand the loss of accepting or rejecting the lot, which directionally decreases with the number of items sampled (as more information of the lot quality is available) and on the other hand the cost of sampling, which increases with this number.

As risk-optimized sequential sampling plans cannot easily be obtained analytically, iterative techniques and algorithms are presented to obtain these plans. Numerical examples of such plans are given, and the results of sensitivity analysis are presented.

CONTENTS

CHAPTER III

Analytical solutions for some k-point prior models

CHAPTER IV

Sensitivity Analysis

CHAPTER V

Conclusions and directions for future work

NOTATIONS

a_i	Prior probability of occurrence of p_i.
$\hat{a}_i$	Posterior probability of occurrence of p_i.
ac	Subscript designating the decision: "accept lot".
ASN	Average Sampling Number of the plan.
c	Cost of sampling one item.
ceil(x)	The smallest integer greater than or equal to x.
cn	Subscript designating the decision: "continue sampling".
δ	Loss due to imperfect information.
floor(x)	The largest integer lesser than or equal to x.
k	the number of components of the discrete k-point prior distribution.
LTPD	Lot Tolerance Percent Defective.
n	Number of items sampled.
n_m	Cumulative number of items sampled sequentially to reach the meeting point.
(n,r)	Coordinates of a point in the sample space.
(n_m, r_m)	Coordinates of the meeting point in the sample space.
N	*ceil*(n_m).
N^*	The largest possible sample size (the effective meeting point) for a plan.
ϕ	The probability of an inspection station making a Type I error (rejection of a non-defective item).
$\mathbb{P}$	The sampling path, a locus of (n,r) points as sampling proceeds.

p	The true unknown lot quality, the measure of the fraction of defective items in a lot.
$\overline{p}$	The apparent lot quality as observed (will in general be different from p due to imperfect inspection).
p_i	The i-th probable value of the distribution of the lot quality p.
q_{rj}	Probability of finding a defective item at the current stage of sampling.
r	The "rejection count" — number of items found defective in sequentially sampling n items.
r_m	Cumulative number of defective items found when the meeting point is reached.
$r_{ac}(n)$	The rejection count at the acceptance boundary at n.
$r_{rj}(n)$	The rejection count at the rejection boundary at n.
rj	Subscript designating the decision: "reject lot".
$R_{xy}(n,r)$	The risk of making decision xy $\left[\in \left\{ac,\ rj,\ cn\right\}\right]$ at a point (n,r) on the sample space.
$R^*(n,r)$	The optimum Bayes' risk at a point (n,r) on the sample space.
θ	The probability of an inspection station making a Type II error (acception of a defective item).
$W_{ac}(p)$	Loss caused by accepting the lot at lot quality p.
$W_A(p)$	(See $W_{ac}(p)$.)
$W_{rj}(p)$	Loss caused by rejecting the lot at lot quality p.
$W_R(p)$	(See $W_{rj}(p)$.)

GLOSSARY†

AQL (Acceptable Quality Level)
(ASQC Standard)
The AQL is the maximum percent defective that for the purposes of sampling inspection can be considered satisfactory as a process average.

ASN
Average Sampling Number. The average number of items sampled in using a sampling plan.

BAYES' RISK (synonym for *EXPECTED RISK*)

DECISION BOUNDARY
The boundary on the *sample space* which separates the "continue sampling" region from one of the *terminal decision* regions.

EXPECTED RISK (of the sampling plan)
The expected consequence of operating the sampling plan in terms of *loss* of the (expected) *terminal decision*. Called Bayes' Risk in literature. (See *LOSS*.)

LOSS (caused by a decision)
A quantified measure of all unfavorable consequences of a decision.

LOT QUALITY

The fraction of defective items in a lot.

LTPD (Lot Tolerance Percent Defective)

The percent defective at which the consumer's risk is specified. The LTPD is considered to be a non-acceptable process average.

MEETING POINT

The point on the *sample space* where two *decision boundaries* meet. (Wetherill [19]).

OC CURVE (Operating Characteristic Curve)

The relationship between the probability of acceptance of the lot (based on a sampling plan) and the *lot quality*.

OSP (Optimal Sampling Plan)

A sampling plan obtained by minimizing the *expected* (or *Bayes'*) *risk*.

PRIOR & POSTERIOR PROBABILITIES

Prior probabilities are the probabilities assigned directly to the values of an unknown (random) variable before some particular sample is taken. Posterior probabilities are the probabilities as revised in the light of the additional information from that sample. Note: this distinction between prior and posterior is always *relative* to some particular sample.

RISK

Usually used interchangeably with *loss*. (See also *EXPECTED RISK*.)

SAMPLE SPACE

The space on which the process of *sequential sampling* is represented. Usually it is an integer space with the x-axis marked by the "number of items sampled" and the y-axis marked by "number of items found defective".

SAMPLING

The process of randomly selecting and inspecting items from a homogeneous lot.

SEQUENTIAL SAMPLING

The process of sampling *and* inspecting one item at a time until a *terminal decision* is reached.

TERMINAL DECISION

Any of the decisions with which *sampling* terminates. Usually, "accept lot" and "reject lot" are the only two terminal decisions.

TYPE I ERROR

Rejection of the null hypothesis when it is actually true.

TYPE II ERROR

Acceptance of the null hypothesis when it is actually false.

CHAPTER I

DECISION THEORY AND SEQUENTIAL ACCEPTANCE SAMPLING

1.0 INTRODUCTION

The theoretical foundation of (non-Bayesian) statistical decision theory was laid by Abraham Wald [16], who used minimax procedures for minimizing the maximum loss of a decision. Wald also first introduced the concept of using decision theory in finding solutions to sequential decision problems. We will here be concerned with an alternative proposition for making sequential decisions known as the *Bayesian* statistical decision theory as explained by Pratt, Raiffa and Schlaifer [13].

In this chapter we introduce the problem of constructing optimal sequential sampling plans in the conceptual framework of statistical decision theory as originally done by Barnard [2] and Wetherill [19]. Beginning with an introduction to sequential decisions, we proceed to a statement of the problem and show how some of these plans would appear in practice. We review the salient results obtained by earlier researchers working on this problem to bring out some of the important issues in developing optimal sequential sampling plans. We then present an overview of the present work, as developed in Chapters II and III.

1.1 WHAT ARE SEQUENTIAL DECISIONS?

In sequential decision making, statistical decision theory appears to be the natural approach for guiding the decision maker toward "optimal" decisions. This may be illustrated as follows. Suppose that a decision maker is faced with the problem of selecting a decision d from a set of possible decisions D. The effectiveness of the decision he makes is dependent upon an unknown parameter p (e.g., the quality of the lot being appraised), information about which can be "purchased" at a cost of c per "unit of information". Suppose that the risk in making the decision d when the parameter is p is $W_d(p)$. The rational decision maker wishes to minimize the total risk of his making the decision, which includes the total cost of information purchased in making the decision. If he employs the Bayesian approach, the decision maker may start with some *a priori* knowledge about the parameter p, and at any stage he has the choice of either making a "terminal" decision $d \in D$, or of deferring the terminal decision until more information about p is available. Each purchase of information about the parameter p represents a "stage". The decision maker would proceed sequentially from one stage to the next, evaluate his risk, and terminate the process of decision making when the total risk of making a decision becomes minimum for all stages.

To solve the above problem optimally, i.e., to develop the sampling plan to minimize his total risk it has been suggested that the decision maker could use decision theory to guide him in choosing the best decision at every stage. A large volume of appropriate literature is now available that helps tackle this problem. One may refer to DeGroot [5] for a rigorous theoretical presentation of the methods. According to Bayesian statistical decision theory, the best (i.e. optimal) decision would be the one that minimizes the *expected total risk given the current information about parameter p*. The term "risk" here stands for the loss of utility as explained in Pratt *et al* [13]. As a special case, risk can represent the monetary loss from a decision, provided that monetary loss is a satisfactory indicator of the decision maker's utilities.

The correct assessment of prior information is as important as the effect this information might have on the outcome of the decision making process. If a "large" volume of additional information is purchased during decision making, the effect of prior information will be comparatively low. (How large is "large" will depend on the parameters of the problem.) If, on the other hand, the volume of information purchased is "low", the effect of prior information will be significant. Barnard [2] has argued that the effect of prior information will usually be significant in sequential decision problems because the objective itself is to minimize the amount of information collected, for often the cost of obtaining this information is high.

1.2 OPTIMAL SEQUENTIAL SAMPLING PLANS

Most sequential sampling plans employed in guiding lot acceptance/rejection decisions today are those based on Wald's sequential probability ratio test (SPRT) [15]. These plans are based on *statistical* measures of consumer's and producer's risks (expressed as probabilities or odds) at the AQL and LTPD respectively. SPRTs suffer from two limitations, however. While these plans are very useful in limiting the expected total number of items sampled from a lot, they cannot be called "optimal", because they do not minimize any function of cost or risk in the decision theoretic sense. Also, they do not offer a generalized model for constructing a more accurate representation of the risk to the decision maker; SPRTs consider risks specified only at two points – the AQL and the LTPD. The above two limitations of existing sequential sampling plans can be overcome by constructing plans based on a decision theoretic formulation of the problem.

We first observe that a sequential sampling plan is basically a sequential procedure to (optimally) guide decision making about accepting or rejecting a lot (or batch) of items that has been submitted for acceptance. The two terminal decisions possible here are to "accept" or "reject" the lot keeping in perspective the unknown parameter, p, known as the *lot quality*. p represents the

true fraction of items that are defective in the lot. Note that p is not a measure of the fraction of items that are *found* defective by inspection. In general, these two measures will be different due to the presence of inspection errors. The *a priori* knowledge of the value of p is represented as a prior distribution of p, the unknown parameter. Additional information about p may be obtained by randomly sampling and inspecting one or more items from the lot, and upon inspection classifying them either as defective or non-defective, the cost of each such operation being c. Bayes' theorem is used for updating the prior probabilities based on the additional information obtained from sampling. The risk in accepting or rejecting the lot when lot quality is p is denoted by losses $W_A(p)$ and $W_R(p)$ respectively, the objective of the decision maker always being minimization of the total expected risk. In sequential sampling, after obtaining every sample, the decision maker decides optimally on either a terminal decision, or on continuing sampling based on his knowledge of the lot quality p at that stage of sampling.

The problem of developing a sequential sampling plan is thus similar to the one in Section 1.1 and can be solved optimally.

To visualize what possibly the sampling plan would be like when the above problem is solved, consider a 2-D plane (Figure 1.2.1) with the horizontal axis representing the number of items sampled, n, and the vertical axis representing the cumulative number of items found

defective, r. This description follows the formalisms put forth by Barnard [1], [2] and Wetherill [19]. The (n, r) space may be treated as a representation of the "sample space" of the sampling procedure, with each point in it represented as (n, r). The sequential sampling process can be represented as a path $\mathbb{P}$ on the sample space beginning at (n = 0, r = 0), with each successive point at stage i having the n-coordinate as i, and the r-coordinate as the cumulative number of items found defective after i items have been inspected. The entire sample space may be now divided into three separate regions, each of which indicates the optimal decision to be taken when the path $\mathbb{P}$ enters this region. The three regions correspond to the decisions "accept lot," "reject lot," and "continue sampling," and thus constitute the sampling plan, giving both the procedure and the criteria for making lot acceptance/rejection decisions.

The boundary between the *accept* and *continue* regions is called the "acceptance" boundary and that between the *reject* and *continue* regions the "rejection" boundary. These boundaries are collectively called "decision boundaries". The point at which these three decision boundaries meet is called the "meeting point" (Figure 1.2.1).

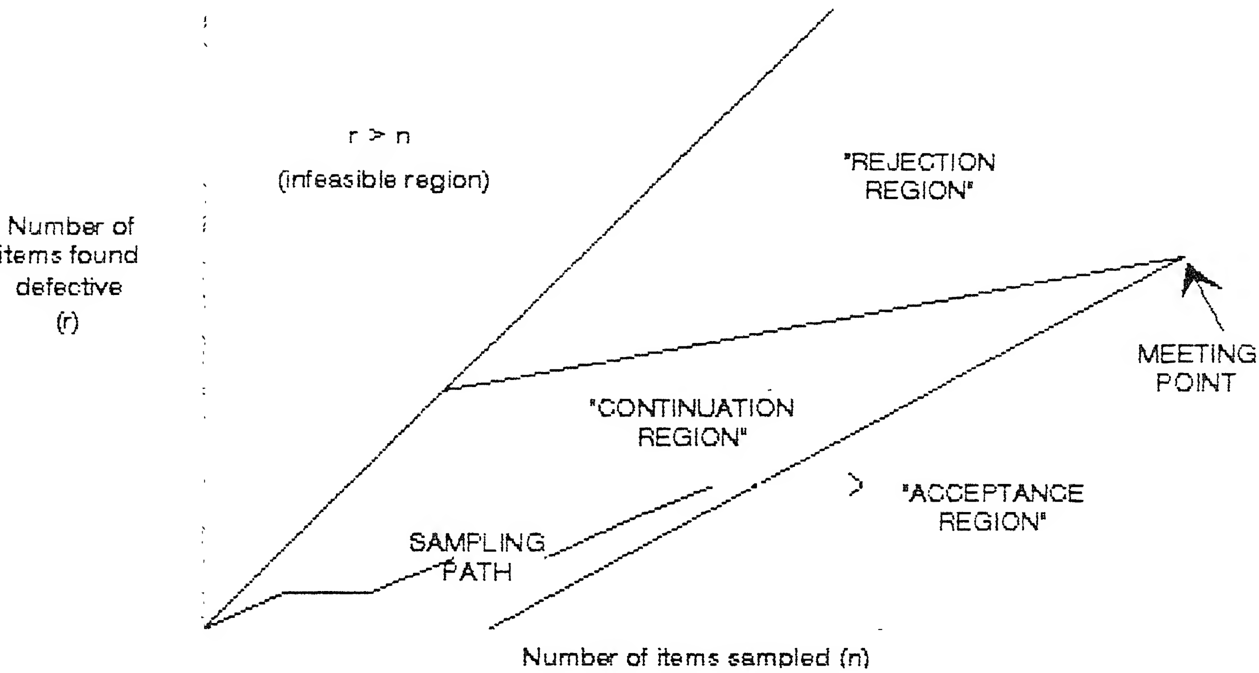

FIGURE 1.2.1. The sequential sampling process shown on the (n, r) Sample Space.

While the procedure for finding the regions for a sequential sampling problem will not be described in detail until Chapter III, it is useful to see how the elements of the sequential sampling plan relate to each other and then lead to the optimal plan.

According to Johnson [9], an optimal sequential sampling plan derived from statistical decision theory comprises the following elements:

(1) Z, The objective of the decision maker:
Usually the minimization of total expected risk.

(2) Terminal decisions, $\{d_i\}$:
Usually "accept lot" and "reject lot".

(3) The Prior distribution $\xi(p)$:
Usually that of the lot quality. This distribution can be specified as either discrete or continuous depending upon the solution methodology.

(4) Risk, $W_D(p)$:
The loss of utility (in the decision theoretic sense) associated with each terminal decision D at lot quality p.

(5) The cost of information, c:
The cost of sampling and inspecting one item.

(6) The optimal sequential sampling plan, S:
This is the resulting optimal sequential sampling plan.

These elements of optimal sequential sampling plans and their inter-dependencies are shown in Figure 1.2.2.

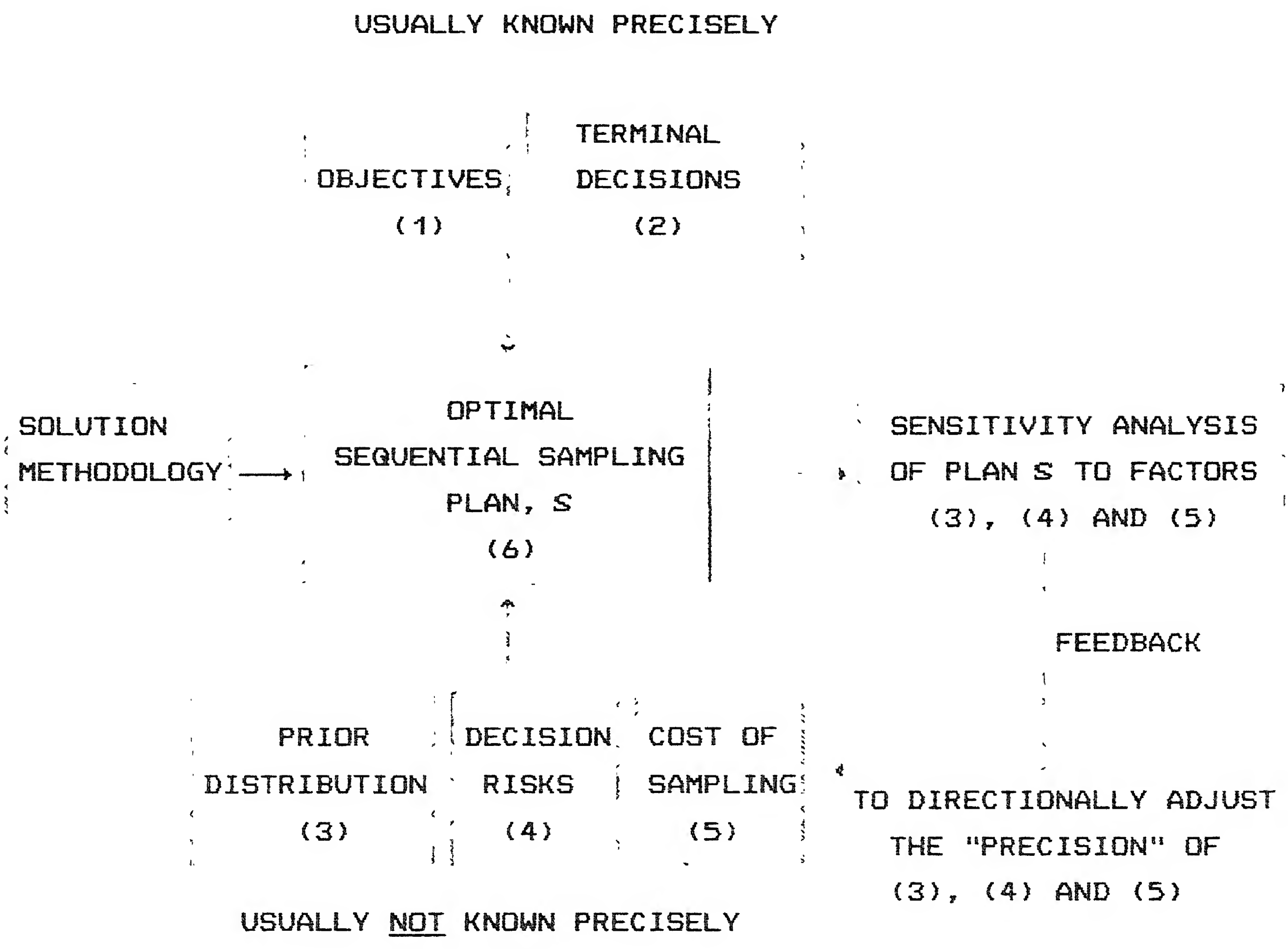

FIGURE 1.2.2. Construction of optimal sequential sampling plans.

A summary of the approaches that variously optimize the sequential decision process in acceptance sampling follows.

1.3 THE METHODS FOR DEVELOPING OPTIMAL SEQUENTIAL SAMPLING PLANS

1.3.1 BARNARD'S LIKELIHOOD RATIO TEST [2]

In this important paper, Barnard pointed out the close correspondence between the theory of statistical decisions and the theory of sampling inspection. Barnard stated that the motivation for constructing sampling plans on the basis of prior information is that neighboring batches of items "resemble" each other in terms of fraction defective, and this information could be used in acceptance sampling.

Barnard stressed the importance of the form chosen for the prior distribution, by demonstrating using an example that if we try to fit the prior distribution by a frequency curve belonging to a *family* of such curves, then this family must involve more than two parameters to avoid inconsistency in the results. It is instructive to note that the commonly used continuous prior distributions, i.e., normal, beta and gamma, all belong to the family of curves having only two parameters, and hence, according to Barnard, cannot be used as a general form for fitting prior distributions. The mixed-binomial form, in the simplest case of the two-point prior, has *three* parameters, and should thus be preferred over any two parameter family. In addition, apart from its theoretical advantage, Barnard pointed out that the mixed-binomial distribution also occurs quite commonly in real life.

When the prior is represented by a mixed binomial with only two components, p_1 and p_2 respectively (which are two possible values of the lot quality of a batch of items, with $p_1 < p_2$), Barnard showed that the "most economical" sampling plan (one that minimizes the total expected risk) is obtained by a likelihood ratio test procedure. If a_1 and a_2 are the prior probabilities of p_1 and p_2 respectively ($a_1 + a_2 = 1$), and if $(1 - p_i)$ is denoted by q_i, the likelihood ratio is defined by

$$l(x,y) = \lambda = a_1\, p_1^x\, q_1^y \,/\, a_2\, p_2^x\, q_2^y \tag{1.3.1.1}$$

where x defectives and y non-defectives have been observed. The risks of errors are given by $R_a(p_2)$ and $R_r(p_1)$, the risks of accepting and rejecting the batch at lot quality p_2 and p_1 respectively. One would accept the batch if

$$l(x,y) = p_1^x\, q_1^y \,/\, p_2^x\, q_2^y \;>\; a_2\lambda_1/a_1 \tag{1.3.1.2}$$

and would reject it if

$$l(x,y) = p_1^x\, q_1^y \,/\, p_2^x\, q_2^y \;<\; a_2\lambda_2/a_1 \tag{1.3.1.3}$$

while if $l(x,y)$ lies between these two limits, i.e. $\lambda_2 < \lambda < \lambda_1$, it would be worthwhile to continue with more inspection. The aim is to find λ_1 and λ_2 such that the total expected risk is minimized. If we denote U by $a_2\lambda_1/a_1$ and L by $a_1/a_2\lambda_2$, the problem is transformed to that of finding U and L optimally. To solve the problem, it is required to estimate $S(U,L,p)$, the mean sample size of the

sequential procedure, and A(U,L,p) and R(U,L,p) the probabilities of acceptance and rejection when the lot quality is p. The exact formulae from which S, A and R can be calculated have been given by Burman [4]. It is then possible to derive the equations that determine the values for λ_1 and λ_2.

1.3.2 VAGHOLKAR AND WETHERILLS' 2-POINT PRIOR MODEL [14].

Based on basic theory given by Barnard [2], Vagholkar and Wetherill have given a method of arriving at the optimum test procedure represented by the acceptance and rejection boundaries which for the two point prior situation do not meet. (A meeting point is desirable in a sequential procedure to rule out the possibility of an arbitrarily large sample size before a decision is made.) The equations obtained for calculating the decision boundaries are similar to those obtained by Barnard (indeed, because the model is the same), and are solved numerically by iteration to find the values for λ_1 and λ_2. The sequential procedure can then be implemented using the scoring procedure due to Barnard [1], obtained by taking logarithms on both sides of Equation (1.3.1.1). Two illustrative numerical examples are also provided by Vagholkar *et al*.

As long as the cost of inspection is linearly related to the number of items sampled, the above sampling scheme will be optimum. It is suggested by these authors that when the cost structure also

includes the cost of administrating sequential sampling schemes and the cost of pre-processing the items before inspection, the optimal scheme may be found by the consideration of group sampling schemes having a fixed group size.

Vagholkar *et al* also state that for multi-point prior distributions with the number of components greater than or equal to three, the decision boundaries will in general meet. We feel, however, that finding the meeting point analytically may not be easy for $k \geq 4$.

1.3.3 WETHERILLS' CONJUGATE FAMILY PRIOR [19, 20].

In this important work, Wetherill has discussed a number of points arising out of the problem of constructing optimum sequential sampling procedures.

Wetherill shows that a great simplification in formulating the sequential sampling problem is introduced by considering prior distributions "closed" under sampling (i.e., distributions whose posteriors belong to the same family as the priors with respect to sampling from a certain density function – such a family is called the "conjugate family"). Suppose that the prior knowledge is represented by the distribution $\xi(\theta|\alpha)$ of the (vector) decision parameter $\theta \in \Omega$, depending on a known (vector) constant α. Let the result of inspection of an item be a random variable X with a density function $\phi(x|\theta)$. After an observation x_i is taken, the

posterior distribution of θ becomes

$$\frac{\xi(\theta|\alpha)\ \phi(x_i|\theta)}{\int_\Omega \xi(\theta|\alpha)\ \phi(x_i|\theta)\ d\theta}$$

If this posterior distribution has the same functional form as that of $\xi(\theta|\alpha)$, say $\xi(\theta|\beta)$, then the prior form ξ is said to possess the property of being closed under sampling from a distribution ϕ. In other words, ξ is said to belong to a conjugate family of prior distributions.

For distributions closed under sampling, the sample space of the sequential sampling scheme is called the ξ-space, the axes of which are represented by the vector α. Thus each member of the family of distributions $\xi(\theta|\alpha)$ is represented by a point on the ξ-space. Wetherill derives the equations which lead to the meeting point on the decision boundaries on the ξ-space, and gives a dynamic programming scheme to find the decision boundaries starting from this meeting point.

The equations leading to the meeting point are obtained as

$$\int \xi(\theta|\alpha)\ [W_r(\theta) - W_a(\theta)]\ d\theta = 0 \qquad (1.3.3.1)$$

and

$$\int q(\alpha,\theta)\ [W_r(\theta) - W_a(\theta)]\ \xi(\theta|\alpha)\ d\theta = 0 \qquad (1.3.3.2)$$

where, $W_a(\theta)$ and $W_r(\theta)$ are the risks of making decisions accept and reject respectively given that the decision parameter is θ. $q(\alpha,\theta)$ is the probability that for an α on the meeting point boundary, one observation leads to points in the acceptance region.

Wetherill proves that both the decision boundaries do in fact meet at the meeting point obtained from Equations (1.3.3.1) and (1.3.3.2). The optimal *terminal* risk at any point β in the ξ-space is given by

$$R_d(\beta) = \min \left\{ \int \xi(\theta|\beta)\, W_1(\theta),\ \int \xi(\theta|\beta)\, W_2(\theta) \right\} \tag{1.3.3.3}$$

while the *continuation* risk can be calculated optimally by

$$R_c(\beta) = 1 + \iint \xi(\theta|\beta)\, \phi(x|\theta)\, R(\beta^+)\, dx\, d\theta \tag{1.3.3.4}$$

where β^+ corresponds to the posterior distribution for a prior distribution β and an observation x. The optimal risk at β is then

$$R(\beta) = \min \left\{ R_c(\beta),\ R_d(\beta) \right\} \tag{1.3.3.5}$$

Now, if point β is reached after n sampling steps, the risk $R_c(\beta)$ cannot be calculated unless the optimal risks of all points β^+ reached in (n + 1) steps is known. If N is the sample size corresponding to the meeting point, then decisions at all points reached in N steps (one such point is the meeting point) have to be terminal decisions, and the minimum risk at each such point can be

calculated easily as in Equation (1.3.3.3). Thus it is possible to locate the meeting point at (N, R) by solving Equations (1.3.3.1) and (1.3.3.2) simultaneously, and then work backwards in n to find the optimal risks of all points corresponding to sample size (N−1), (N−2), ..., n, ..., 0, using the Equations (1.3.3.4) and (1.3.3.5). The above principle gives an optimal sampling plan, because it conforms to the principle of optimality given by Bellman [3].

1.4 THE PRESENT WORK: THE K-POINT PRIOR WITH K ≥ 2.

In this work, we find optimal sequential sampling plans based on the general k-point prior distribution with $k \geq 2$ using the concept of the meeting point developed by Wetherill [19], and thus extend (1) the range of priors available for developing sequential plans by including the k-point prior distributions, and, (2) the ability to work with prior distributions with the number of parameters greater than 2.

We study the following issues in the construction of optimal sequential sampling plans using the k-point form of the prior distribution:

1) Formulation of the problem of obtaining optimal sequential sampling plans and the derivation of equations for obtaining the decision boundaries.

2) Computational aspects of the construction of the decision boundaries for the optimal sequential sampling plans.

3) Calculation of the ASN and the OC curves for the k-point prior case.

4) Calculation of the expected risk in using the sampling plan.

5) Sensitivity of the plan parameters (e.g., risk, ASN, Pa) to the following factors:

    1) The cost of inspection.
    2) The parameters of the prior distribution
    3) Location of the meeting point.
    4) Imperfect knowledge of the inspection cost
    5) Imperfect knowledge of the decision losses
    6) Imperfect inspection.

6) Exact analytical solution for the meeting point for the 2- and 3-point prior case. No meeting point is found for the k = 2 case but a solution is obtained by forcing a meeting point at a sufficiently large n. This is then an alternate way to approach the Vagholkar *et al* [14] solution based on the likelihood ratio test.

CHAPTER II

FORMULATION OF OPTIMAL SEQUENTIAL SAMPLING PLANS BASED ON THE K-POINT PRIOR DISTRIBUTION

2.0 INTRODUCTION

The most popular *continuous* prior distributions used in the literature for deriving optimal sequential sampling schemes have been the beta and the gamma distributions, both specifiable completely by two parameters. Lindley and Barnett [12] have given sequential plans and also numerical results for the beta prior distribution, sampling being done from a binomial population, assuming the losses of the terminal decisions to be linear. Plans for the gamma-Poisson case have been given by Lechner [11]. The *discrete* prior distribution for which analytical results are available is the 2-point prior distribution. Vagholkar *et al* [14] derive the equations for the 2-point prior case using the likelihood ratio test given by Barnard [2], and outline an iterative procedure for obtaining the decision boundaries.

In the present work, we derive optimal sequential sampling schemes using the generalized k-point prior distribution assuming arbitrary loss functions. We use the meeting point approach suggested by Wetherill [19] to find the decision boundaries.

As pointed out in the literature, a k-point prior distribution offers the following advantages over the more commonly used conjugate prior distributions:

1) The k-point prior is a less restrictive form for representing prior knowledge. Other conjugate prior distributions (such as beta, gamma and normal) have only two parameters and have been argued against by Barnard [2] who showed that prior distributions based on only two parameters might give inconsistent results and commented that such distributions are not a *general* representation of prior knowledge. The k-point distribution has (2k − 1) parameters that may be independently adjusted to best represent prior information.

2) Barnard [2] also points out that in real life, the distribution of fraction defective in a lot will usually be a multi-point distribution, especially when the lot comes from a process operating at several different levels of quality.

For $k = 3$, the resulting model can be solved analytically and exactly to produce the meeting point. As we show, for $k \geq 4$, the model can be solved numerically to converge to the exact location of the meeting point of the acceptance and rejection boundaries. We find, however, that the numerical procedure does not always converge to a solution (if it exists). This may require further study. If

the meeting point is forced at a sufficiently large N, the optimum decision boundaries can be found optimally even if the meeting point cannot be found analytically or numerically.

2.1 FORMULATION OF THE k-POINT PRIOR MODEL

In the k-point prior model, it is assumed that the unknown lot quality (fraction defective) p can have k distinct values p_1, p_2, ..., p_k . With each possible p_i, the decision maker associates a prior probability of its occurrence, a_i, based on the knowledge existing about the possible distribution of lot quality. Thus the k-point prior distribution P of the lot quality p becomes

$$P(p = p_i) = a_i \quad \text{for } i = 1, 2, \ldots, k, \tag{2.1.1}$$

where $\sum_{i=1}^{i=k} a_i = 1$, $k \geq 2$. (Figure 2.1.1)

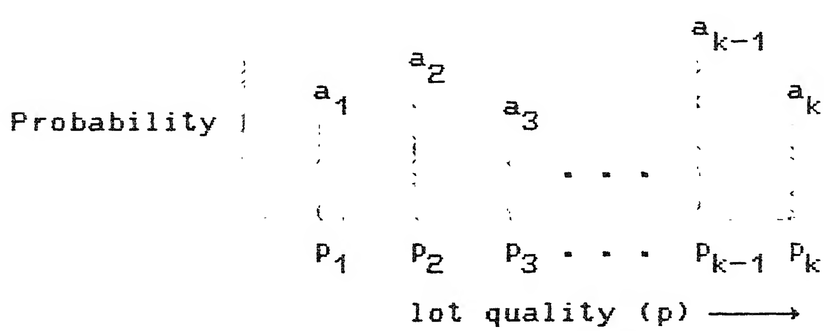

FIGURE 2.1.1. The k-point prior distribution.

Sequential sampling to guide acceptance/rejection of the lot would proceed by sampling and inspecting one item at a time at every stage of sampling, the cost of this operation being c each time an item is inspected (c is assumed to be known to the decision maker). The decision at each stage would be guided by a sampling plan, which would be represented by the acceptance, rejection and continuation regions on the sample space as discussed in Section 1.2. Let the three decisions possible at every stage, namely "accept lot", "reject lot", and "continue sampling", be abbreviated by notations "ac", "rj" and "cn" respectively. The losses associated with each of these decisions for any given lot quality p may be denoted by $W_{ac}(p)$, $W_{rj}(p)$ and $W_{cn}(p)$ respectively and are assumed to be known to the decision maker. The objective of the decision maker is to sample items one by one to eventually accept ar reject the lot such that the total (expected, Bayes') risk is minimized.

We will attempt to find the optimal decision regions for the above sequential decision problem following the procedure given by Wetherill [19]. The procedure begins by locating the meetinq point of the decision boundaries, and subsequently uses an iterative scheme (a recursive scheme converted into an iterative scheme) to find the three decision regions for all sample sizes less than or equal to the maximum possible sample size (the maximum sample size corresponds to the sample size required to reach the meeting point). This procedure results in optimal decisions (decisions which minimize the total expected risk) if the meeting point can be accurately located (Wetherill [19]).

2.2 EQUATIONS FOR DETERMINING THE MEETING POINT

At the meeting point (n_m, r_m) of the decision boundaries, by definition, the risk associated with sampling one more item is equal to the risk of making either of the two terminal decisions, namely, accept or reject the lot. Also, the risk of continuing with sampling is equal to the cost of taking the next sample plus the risk of making an optimal terminal decision subsequently. If $q_{rj}(p)$ denotes the probability of rejecting the lot at the meeting point, the risk of continuation at the meeting point is

$$R_c(p) = q_{rj}(p)\, W_{rj}(p) + (1 - q_{rj}(p))\, W_{ac}(p) \tag{2.2.1}$$

The *total expected risk* of continuation R_{cn} at the meeting point is found by "weighting" $R_c(p)$ with the k-point posterior distribution at (n_m, r_m), and adding the cost of sampling c,

$$R_{cn} = c + \sum_{i=1}^{i=k} \left[q_{rj}(p_i)\, W_{rj}(p_i) + (1 - q_{rj}(p_i))\, W_{ac}(p_i) \right] \hat{a}_i(n_m, r_m) \tag{2.2.2}$$

$$\text{where} \quad \hat{a}_i(n_m, r_m) = \frac{a_i\, (p_i)^{r_m}\, (1 - p_i)^{n_m - r_m}}{\sum_{i=1}^{i=k} a_i\, (p_i)^{r_m}\, (1 - p_i)^{n_m - r_m}} \tag{2.2.3}$$

is the posterior distribution at (n_m, r_m), obtained by applying Bayes Theorem, having started the sampling process with the prior (2.1.1). (*cf* page 282, Wetherill [20].)

The expected risks of the terminal decisions accept and reject at the meeting point are given by

$$R_{ac} = \sum_{i=1}^{i=k} W_{ac}(p_i)\ \hat{a}_i(n_m, r_m) \quad \text{and} \quad R_{rj} = \sum_{i=1}^{i=k} W_{rj}(p_i)\ \hat{a}_i(n_m, r_m) \qquad (2.2.4)$$

At the meeting point, $R_{ac} = R_{rj} = R_{cn}$, and $q_{rj}(p) = p$. Equating R_{ac} and R_{cn}, and substituting for $q_{rj}(p)$,

$$c + \sum_{i=1}^{i=k} \left[p_i W_{rj}(p_i) + (1 - p_i) W_{ac}(p_i) \right] \hat{a}_i(n_m, r_m) = \sum_{i=1}^{i=k} W_{ac}(p_i)\ \hat{a}_i(n_m, r_m)$$

or,

$$\sum_{i=1}^{i=k} \left[W_{ac}(p_i) - W_{rj}(p_i) - c/p_i \right] \hat{a}_i(n_m, r_m) = 0 \qquad (2.2.5)$$

Because we are at the meeting point, we can equate R_{ac} and R_{rj}, resulting in

$$\sum_{i=1}^{i=k} \left[W_{ac}(p_i) - W_{r}(p_i) \right] \hat{a}_i(n_m, r_m) = 0 \qquad (2.2.6)$$

Substituting for $\hat{a}_i$ in equations (2.2.5) and (2.2.6), we get

$$\sum_{i=1}^{i=k}\left[W_{ac}(p_i) - W_{rj}(p_i) - c/p_i \right] p_i a_i (p_i)^{r_m} (1 - p_i)^{n_m - r_m} = 0 \quad (2.2.7)$$

$$\sum_{i=1}^{i=k}\left[W_{ac}(p_i) - W_{rj}(p_i)\right] a_i (p_i)^{r_m} (1 - p_i)^{n_m - r_m} = 0 \quad (2.2.8)$$

Equations (2.2.7) and (2.2.8) are two simultaneous non-linear equations in two variables (n_m, r_m) — the coordinates of the meeting point. If the meeting point exists (conditions not yet made clear in the literature), solving these equations will in general be possible numerically using any of the various quardrature methods available for solving such equations in two variables. The Newton-Raphson procedure is one possible approach which could be used here: we have found that it is relatively simple to implement and converges quickly. For the description of this procedure, one is referred to Kelly [10], or to any other standard text on numerical methods. Analytical solutions are possible for small values of k – the exact solutions for k = 2 and k = 3 are presented in Chapter II.

2.2.1 INTERPRETATION OF THE MEETING POINT

It is necessary here to clarify the concept of the "meeting point" in the context of the sample space (Section 1.2). We will use here the following standard notations:

ceil(x) : the smallest integer greater than or equal to x, and,

floor(x): the largest integer less than or equal to x,

where x is a real number.

The sample space is a discrete space, with both n and r taking only integer values. The solution of Equations (2.2.7) and (2.2.8), if it exists, in general will yield non-integer (real) values for n_m and r_m. Hence the "real" meeting point (n_m, r_m) is not realizable during sampling. But these (real) values of n_m and r_m give us two trivial, but important results:

1) Acceptance is the optimum decision at all points (*ceil*(n_m),r_a) and rejection is the optimum decision at all points (*ceil*(n_m), r_r), where $r_a \leq$ *floor*(r_m) and $r_r \geq$ *ceil*(r_m) (see Figure 2.2.1.1).

2) The maximum possible sample size N is *ceil*(n_m), beyond which only terminal decisions (i.e, "accept" and "reject", but <u>not</u> "continue" sampling) are optimal.

(N, *ceil*(r_m)): "REJECTION" OPT

(n_m, r_m)   (N, *floor*(r_m)): "ACCEPTANCE" OF

FIGURE 2.2.1.1. Interpretation of the meeting point.

The above two results will be used in finding the optimal d regions which minimize Bayes' risk in the following s

Henceforth, we will denote *ceil*(n_m), the sample size at the point, by N, the maximum sample size obtainable.

2.3 CONSTRUCTION OF THE DECISION REGIONS

In this section we derive an iterative procedure fo construction of the decision regions for an optimal seq sampling plan. At any point (n, r) on the sample space, the of the three possible decisions are given by $R_{ac}(n,r)$, $R_{rj}(n,$ $R_{cn}(n,r)$ where subscripts ac, rj and cn have their usual me The optimum risk at this point $R^*(n,r)$ is given by

$$R^*(n,r) = \min \{ R_{ac}(n,r),\ R_{rj}(n,r),\ R_{cn}(n,r) \}$$

ere $R_{ac}(n,r)$ and $R_{rj}(n,r)$ can be calculated as

$$_c(n,r) = \sum_{i=1}^{i=k} W_{ac}(p_i)\ \hat{a}_i(n,r) \quad \text{and}$$

$$_j(n,r) = \sum_{i=1}^{i=k} W_{rj}(p_i)\ \hat{a}_i(n,r) \qquad (2.$$

d $R_{cn}(n,r)$ is obtained by the recursive relation

$$_n(n,r) = c + \Big\{ q_{rj}(n,r)\ R^*(n+1,\ r+1) + (1 - q_{rj}(n,r))\ R^*(n+1,\ r$$

(2.

ere $q_{rj}(n,r)$ is the probability that the next (n+1-th)
mpled will be found defective and hence the lot rejected at
age of sampling.

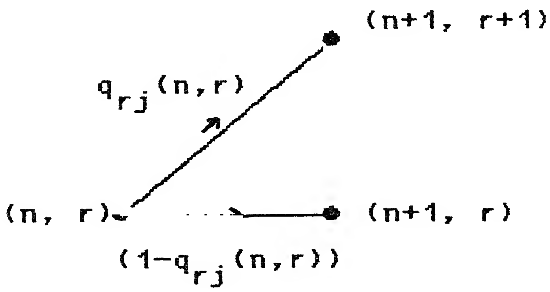

FIGURE 2.3.1. The result of inspecting one item at (n, r)

defective item will lead to the point (n+1,r+1) on the sa
ace while a non-defective item will lead to the point (n+1

Thus if the decision taken after taking one more sample at (n,r) is an optimal decision, the risk $R_{cn}(n,r)$ is obtained as in Equation (2.3.3). In this Equation (2.3.3), $q_{rj}(n,r)$ is calculated from the posterior distribution as

$$q_{rj}(n,r) = \sum_{i=1}^{i=k} p_i \ \hat{a}_i(n, r) \tag{2.3.4}$$

The optimal decisions for all r at N are terminal decisions, the risks of which can be calculated from the Equation (2.3.2). Thus

$$R^*(N,r) = \begin{cases} R_{ac}(N, r) & \text{for } r \le \mathit{floor}(r_m) \\ R_{rj}(N, r) & \text{for } r \ge \mathit{ceil}(r_m) \end{cases} \tag{2.3.5}$$

A recursive scheme based on Equation (2.3.3) will terminate at the N-th stage, where $R^*(N,r)$ will be obtained from equation (2.3.5) without further recursion. Every (n, r) point in the space will thus become marked as an acceptance, reiection, or a continuation point, depending upon which risk is found to be minimum.

2.3.1 SOME COMPUTATIONAL ASPECTS OF THE CONSTRUCTION OF DECISION REGIONS

2.3.1a Conversion of the recursive scheme to an iterative scheme.

The recursion scheme (2.3.3) starting at ($n = 0$, $r = 0$) will require a large amount of storage (in the form of a recursion stack) and will usually be slower. Because it is known beforehand that this scheme will terminate after exactly N steps, it can be converted into an iterative scheme. This iterative scheme begins with calculation of the optimal risks at sample size N from Equation (2.3.5) and proceeds *backwards* in n to find the optimum risks at all the sample points with $0 \leq n \leq N$.

2.3.1b Computational effort.

The most expensive computational element in the above scheme is the calculation of the posterior distribution at every point (n,r). Enumeration of risks at all points with $n \leq N$ in the sample space will require that the posterior distribution be calculated at $(N+1)(N+2)/2-1$ points (as points with $r > n$ cannot be reached). But it is possible to restrict this enumeration to far fewer points in practice because points "beyond" the decision boundaries will never be reached during the sampling process (sampling will terminate with a decision as soon as a boundary is reached). If w(n) denotes the number of points in the continuation region for any n, and the average "width" of the continuation region (w(n) averaged over all

$n \le N$) is $\bar{w}$, the number of points at which the risks have to be enumerated are $(\bar{w} + 2)N$, which is the sum of all continuation and decision boundary points.

(We have found $\bar{w}$ to be usually a small number less than 10 in most problems. In such cases, the posterior needs to be calculated at $O(N)$ points instead of $O(N^2)$ points. $\bar{w}$ is largely a function of the magnitude of the risks, and is large when the risks are large.)

2.3.1c THE LARGEST POSSIBLE "EFFECTIVE" SAMPLE SIZE, N^*.

Sequential sampling will always terminate at or before a certain sample size N^*, defined as the minimum sample size for which there is no continuation point at (N^*+1). In general, N^* will be less than or equal to N ($N = ceil(n_m)$), and we illustrate this using an hypothetical example of decision boundaries shown in Figure 2.3.1.1. Assume that the decision boundaries can be approximated to continuous curves (shown as straight lines in the figure) which separate the acceptance, the rejection and the continuation region. In the illustration, in the vicinity of the (exact) meeting point, the decision boundaries are relatively "narrow" and do not contain a meeting point between them at sample sizes (N–1), (N–3), (N–4) and (N–7). However, there are continuation points at all sample sizes less than (N–7). Thus, (N–7) is the maximum obtainable sample size, N^*, for the sequential sampling plan represented by boundaries shown in the figure. There will, in general, always be a continuation

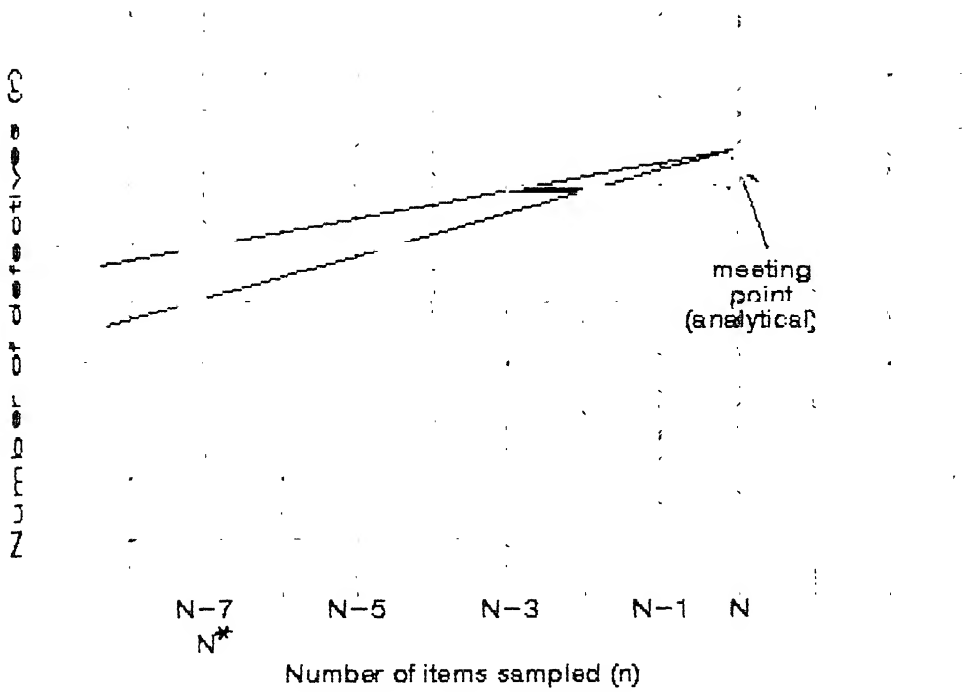

FIGURE 2.3.1.1. The effective sample size, N^*.

point between the two decision boundaries if the separation between them is greater than one unit on the vertical scale.

2.4 "BOUNDS" : THE ALGORITHM TO COMPUTE THE DECISION BOUNDARIES.

Algorithm BOUNDS, given below may be used to compute the decision boundaries for a k-point prior model. This algorithm is implemented in C as shown in Appendix A. Taking the analytical meeting point, the losses, and the prior distribution as the input, BOUNDS

calculates the effective maximum sample size, N^*, and the acceptance and rejection boundaries on the decision space.

The inputs to BOUNDS are:

1) The meeting point, (N, R), where $N=ceil(n_m)$ and $R=floor(r_m)$. Recall that (n_m, r_m) is the analytical meeting point.
2) The risks of accepting the lot $W_{ac}(p)$ and rejecting the lot $W_{rj}(p)$ as functions of the lot quality p.
3) The prior distribution of p (Equation 2.1.1).
4) The cost of sampling per item, c.

The output of BOUNDS are:

1) The effective sample size, N^*.
2) Acceptance and rejection boundaries for all sample sizes $n \leq N^*$, which are obtained as arrays $D_A[n]$ and $D_R[n]$ respectively.

```
algorithm BOUNDS
BEGIN
N* ← N;                    /* initialize effective sample size to N */
MAX_R ← min(N, R + 1);     /* upper bound on the cumulative number of
                              defectives for sample sizes less than N */
D_A[N] ← MAX_R-1;          /* acceptance boundary at N */
D_R[N] ← MAX_R;            /* rejection boundary at N */

    for r ← 0 to D_A[N] do
                     i=k
        R [N, r] ←   Σ  W_ac(p_i) posterior(N, r, i);
                     i=1
    end
    /* above calculates the acceptance risks (optimum risks) at all
    acceptance points at N */

    for r ← R + 1 to R + 2 do
                        i=k
        R [N, R+1] ←    Σ  W_rj(p_i) posterior(N, R+1, i);
                        i=1
    end
    /* calculate the risk of rejection (optimum risk) at (N, R+1)
    and (N, R+2) */

/* backward calculation of risks begins henceforth */
```

for $n \leftarrow N - 1$ to 0 do /* n indexes the sample size */

$d_{ac} \leftarrow 0$

$d_{cn} \leftarrow 0$

for $r \leftarrow 0$ to MAX_R do /* r indexes the number of defectives */

$$R_{ac} \leftarrow \sum_{i=1}^{i=k} W_{ac}(p_i) \text{ posterior\_prob } (n, r, i);$$

/* acceptance risk at (n,r) */

$$R_{rj} \leftarrow \sum_{i=1}^{i=k} W_{rj}(p_i) \text{ posterior\_prob } (n, r, i);$$

/* rejection risk at (n,r) */

$$R_{cn} \leftarrow c + \sum_{i=1}^{i=k} \Big[p_i \; R[s+1][t+1] + q_i \; R[s+1][t]\Big] \times \text{ posterior\_prob } (n, r, i);$$

/* the continuation risk at (n,r). c is the cost of sampling and $q_i = (1 - p_i)$ */

$$R\,[n, r] \leftarrow \min \left\{ R_{ac}, R_{rj}, R_{cn} \right\}$$

/* the minimum risk at (n, r) */

if $R\,[n, r] = R_{ac}$ then $d_{ac} \leftarrow d_{ac} + 1$;

/* d_{ac} counts acceptance points at n */

if $R\,[n, r] = R_{cn}$ then $d_{cn} \leftarrow d_{cn} + 1$;

/* d_{cn} counts continuation points at n */

end /* the loop for r */

```
ʮ[n] ← d_ac;
ʀ[n] ← d_ac + d_cn + 1;
* calculate the acceptance and rejection boundaries at n */

AX_R ← D_R[n];  /* new upper bound on the cumulative number of
                   defectives for sample sizes less than n */

f d_cn = 0 then N* ← n;
  /* if there are no continuation points at n,
  it becomes the new meeting point. See Section 2.3.1c */

nd   /* of the n loop */

ND.  /* of algorithm BOUNDARIES */
```

?.5 THE ASN AND THE OC CURVES

ʮlgorithm "BOUNDS" calculates the decision regions for an optimal sequential sampling plan. We will now discuss the procedure for calculating the average sampling number (ASN) and the operating characteristic (OC) curves for the present sampling plan. The ASN is defined as the expected number of items that are sampled before a terminal decision is reached. The OC curve gives the probability of accepting the lot at any given lot quality p. As we shall see, it is possible to numerically compute the ASN and the OC curves using our procedure *because* the continuation region for the sampling plan is bounded as sampling would not need to beyond the maximum effective sample size N^*.

2.5.1 "REACHABLE" AND "NON-REACHABLE" POINTS

Points *on the decision boundaries* belong to two mutually exclusive (and exhaustive) sets: one consists of "reachable points" and the other of "non-reachable points". All reachable points on the decision boundaries are necessarily termination points. Non-reachable points are so called because they cannot be reached during any sampling process, being "blocked from reach" by other decision boundary points.

We will illustrate the concept of "reachability" and "non-reachability" of points with the help of an example. Let $(n, r_{ac}(n))$ be a point on the acceptance boundary and $(n, r_{rj}(n))$ be a point on the rejection boundary at some stage of sampling n. In Figure 2.5.1, the acceptance point $r_{ac}(6)$ is a non-reachable point because it can only be reached from points (5,0) and (5,1), both of which are points at which acceptance will be preferred, and sampling will stop (in fact, point (5,0) will never be reached). The acceptance point $r_{ac}(7)$ is non-reachable because it can only be reached from (6,0), which lies in the acceptance region, and (6,1), a non-reachable point. Similar arguments can be given for points $r_{rj}(9)$ and $r_{rj}(10)$.

The criteria for a point to be non-reachable is:

1) $r_{ac}(n)$ is a non-reachable point iff $r_{ac}(n) = r_{ac}(n-1)$.
2) $r_{rj}(n)$ is a non-reachable point iff $r_{rj}(n) > r_{rj}(n-1)$.

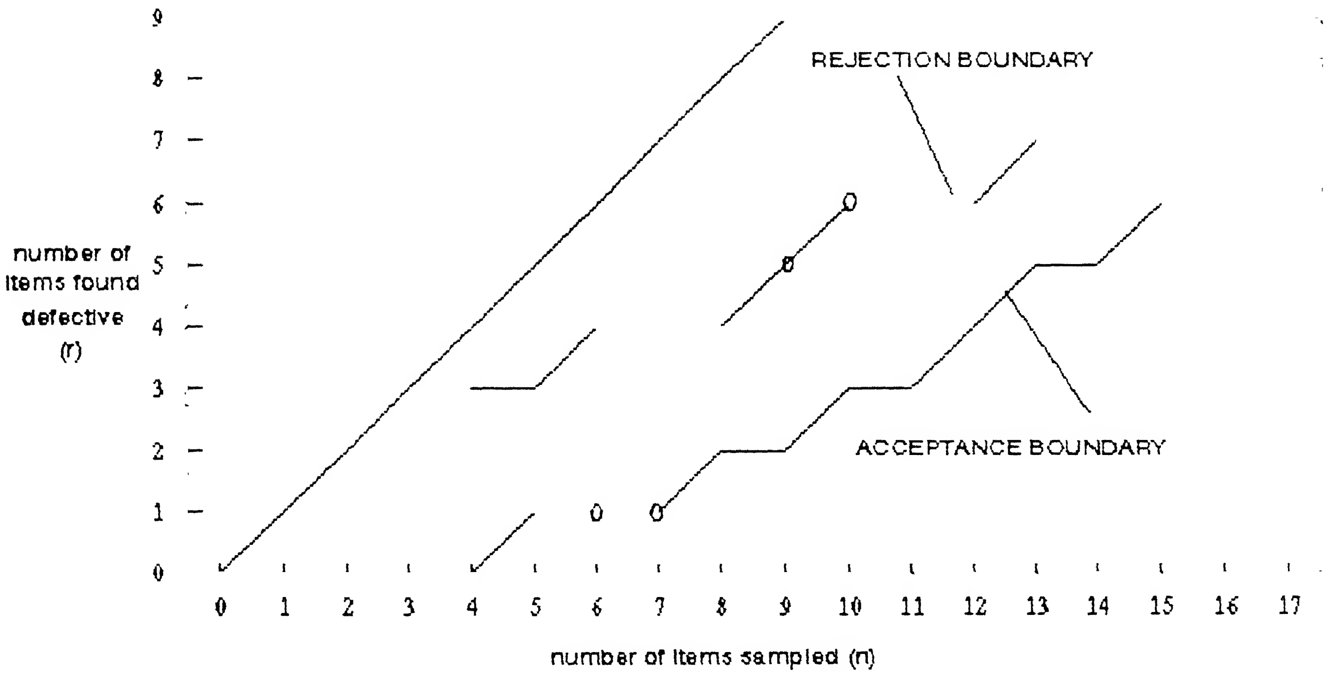

FIGURE 2.5.1. Reachable and non-reachable points.

2.5.2 PROCEDURE FOR THE CALCULATION OF OC AND ASN CURVES

A sampling process can terminate only at a decision boundary. If $P_R(n,r)$ denotes the probability of reaching any point (n,r) on the sample space with lot quality p (Equation 2.5.2.5), then the probabilities of termination of sampling at a point on the decision boundary for sample size n are $P_R(n, r_{ac}(n))$ and $P_R(n, r_{rj}(n))$, for

a given lot quality p. Then the probability of acceptance of the lot, i.e., the OC curve is given by

$$Pa(p) = \sum_{n=1}^{n=N^*} P_R(n, r_{ac}(n)) \tag{2.5.2.1}$$

and the average number of items sampled at lot quality p is

$$ASN(p) = \sum_{n=1}^{n=N^*} n \left[P_R(n, r_{ac}(n)) + P_R(n, r_{rj}(n)) \right] \tag{2.5.2.2}$$

where it should be noted that if $r_{ac}(n)$ or $r_{rj}(n)$ are non-reachable, we have:

$$\left. \begin{array}{l} P_R(n, r_{ac}(n)) = 0 \\ \text{and} \\ P_R(n, r_{rj}(n)) = 0 \end{array} \right\} \tag{2.5.2.3}$$

We now turn to the determination of P_R's in general. The probability of reaching *any* point (n,r) in the sample space by starting sampling at (0,0) is equal to the sum of the probabilities (each sampling being independent) of following each of the unique paths that lead to and end at that point. The probability of reaching any point in the terminal decision region which is not a boundary point is 0. Only the points in the continuation region and the points on the decision boundaries are potentially reachable. Thus, we can calculate $P_R(n,r)$ for all points ($n \leq N$, $r \leq n$) in the decision region using the following recurrence relation:

$$P_R(n,r) = p\,P_R(n-1,\ r-1) + (1-p)\,P_R(n-1,\ r) \qquad (2.5.2.4)$$

with the following boundary conditions:

$$P_R(0,0) = 1 \qquad (2.5.2.5a)$$

$$P_R(n,r) = 0 \quad \text{for } r \geq r_{rj}(n)+1 \quad \text{and} \quad r \leq r_{ac}(n)-1, \text{ and,} \qquad (2.5.2.5b)$$

$$\left.\begin{array}{l} P_R(n,\ r_{ac}(n)) = 0 \\ \text{and} \\ P_R(n,\ r_{rj}(n)) = 0 \end{array}\right\} \text{ for non-reachable boundary points} \qquad (2.5.2.5c)$$

Sequential sampling will eventually always terminate at a point that lies on one of the decision boundaries. One way of looking at this is the following. Consider the following transition matrix, constructed using the Equation (2.5.2.4) and Conditions (2.5.2.5 a, b, and c), showing the *transition probabilities* on the sample space when *one* item is inspected

to: / *from:*	(0,0)	(0,1)	..	(n,n)	..	(n+1,r)	(n+1,r+1)	..	(N^*+1,y)	
(n,r)	0	0	.0.	0	.0.	1−p	p	.0.	0	ROW 1
$(n,r_{ac}(n))$	0	0	.0.	0	.0.	0	0	.0.	0	ROW 2
$(n,r_{rj}(n))$	0	0	.0.	0	.0.	0	0	.0.	0	ROW 3
(N^*,x)	0	0	.0.	0	.0.	0	0	.0.	0	ROW 4

FIGURE 2.5.2.1. Transition probabilities on the sample space.

The transition probability from any boundary point (ROWS 2 & 3) is always zero, and sampling will thus stop at a boundary point. Also, no transition is possible once the effective sample size N^* is reached (ROW 4). Because a "reverse" transition (i.e., $(n,x) \longrightarrow (n-1,y)$) occurs with a zero probability, sequential sampling beginning at (0,0) will result in a (strictly) monotonically increasing sample size and will terminate *in the worst case* at the effective sample size N^* (from which no transition is possible), if a decision boundary point is never encountered during the course of sampling.

2.6 EXPECTED RISK OF THE OPTIMIZED SEQUENTIAL SAMPLING PLAN

The "ideal" decision (based on perfect knowledge of p) will always have a risk that equals the minimum of $W_{ac}(p)$ and $W_{rj}(p)$ (see Section 2.1) at all values of the lot quality p. The optimum risk of such a decision is denoted by

$$R^*(p) = \min \left\{ W_{ac}(p),\ W_{rj}(p) \right\} \tag{2.6.1}$$

Where acceptance will be preferred when $R^*(p) = W_{ac}(p)$ and rejected when $R^*(p) = W_{ac}(p)$.

We assume that it is not economically feasible to make the ideal decision because p is an unknown and the cost of obtaining

information about p may be high. The "economically optimum" sampling plans obtained using the results of the previous sections will in general have an expected risk $R(p) \geq R^{*}(p)$. The expression for $R(p)$ can be obtained using $Pa(p)$ as follows:

$$R(p) = Pa(p)\, W_{ac}(p) + (1 - Pa)\, W_{rj}(p) \tag{2.6.2}$$

with $Pa(p)$ being calculated from Equation (2.5.2.1).

The shift of $R(p)$ from the ideal is the "loss due to imperfect information" and is defined as:

$$\delta = R(p) - R^{*}(p) \tag{2.6.3}$$

It is easily seen that an optimization of $R(p)$ subject to the cost constraints of sampling is equivalent to a minimization of δ, because $R^{*}(p)$ is a constant dependent only on the actual lot quality p. The optimum sequential sampling plan also minimizes δ.

In general, there can be many factors that could affect the optimality of the sampling plan. Some of these factors are imperfect inspection, imprecise knowledge of the various risk and cost parameters, and the subjective differences in the assessment of the prior distribution. The location of the meeting point is another factor which might affect the plan's optimality — when the meeting point cannot be exactly determined or does not exist and may be forced arbitrarily at some large enough sample size to stop the

sampling process. Some of these factors might even lead to a *decrease* in δ as we shall see during sensitivity analysis! The following section is devoted entirely to one such factor, the effect of imperfect inspection (as items are sampled one by one) on the optimality of the plan.

2.7 THE EFFECT OF IMPERFECT INSPECTION

Two types of errors are possible during inspection: rejection of a non-defective item (Type I error) and acceptance of a defective item (Type II error). Studies show that the above two types of error can at times adversely affect the optimality of decision making during acceptance sampling (see Dorris *et al* [6]).

If the probability of the Type I error in inspection is ϕ and that of the Type II error is θ, the fraction of the items that will be classified as "defective" (known as the apparent fraction defective) will be

$$\bar{p} = p(1-\theta) + (1-p)\phi \qquad (2.7.1)$$

If these θ and ϕ errors are not corrected, the decisions regarding the acceptance and rejection of the lot will be based on this apparent lot quality $\bar{p}$ instead of on the true lot quality p. the extra loss incurred due to imperfect inspection will then be

$$\bar{\delta} = R(\bar{p}) - R(p) \tag{2.7.2}$$

If the cost of correcting imperfect inspection from levels (ϕ, θ) to levels $(\hat{\phi}, \hat{\theta})$ is given by $c(\phi, \theta, \hat{\phi}, \hat{\theta})$, the gain resulting from this correction is given by

$$G = R(\bar{p}) - R(p(1-\hat{\theta})+(1-p)\hat{\phi}) - c(\phi, \theta, \hat{\phi}, \hat{\theta}) \tag{2.7.3}$$

When p is known, or can be estimated, a correction $(\phi, \theta) \longrightarrow (\hat{\phi}, \hat{\theta})$ should be applied if $G > 0$.

PROCESS $\xrightarrow{p}$ INSPECTION $\longrightarrow$ $p(1-\theta)+\phi(1-p)$

p fraction defective

$p(1-\theta)+\phi(1-p)$ fraction *called* defective

$p\theta+(1-p)(1-\phi)$ fraction *called* non-defective

FIGURE 2.7.1. The effect of inspection error.

It may be possible to derive the optimal sequential sampling plan given $(\theta\ \phi)$. However, we will not attempt that here.

CHAPTER III

ANALYTICAL SOLUTIONS FOR SOME K-POINT PRIOR MODELS

3.0 INTRODUCTION

While the k-point prior case will in general be usually solved *numerically* to yield the meeting point and the subsequent determination of the decision boundaries, two particular cases, those of $k = 2$ and $k = 3$, can be investigated *analytically*. The discussion of the 2-point prior case can also be found in Wetherill [20] and Vagholkar *et al* [14]. Wetherill shows that the optimal plan for the 2-point prior case is an SPRT, and that the meeting point does not exist when the unit cost of sampling is a constant. Vagholkar *et al* use the likelihood ratio test procedure to find the decision boundaries for the 2-point prior case.

Here we discuss the 2-point prior case using our formulation (which is similar to that of Wetherill [19]) and also present an *exact* analytical solution for the meeting point for the 3-point prior case.

3.1 SOLUTION FOR THE 2-POINT PRIOR DISTRIBUTION CASE

The 2-point prior case is the simplest of all possible k-point distributions. The analytical solution for this case based on the likelihood ratio test of Barnard [2] was give by Vagholkar *et al* [14]. (For the 2-point prior case, it was shown by Barnard that a likelihood ratio test can be constructed to produce the expected minimum cost sampling plans, which he called optimum plans.)

Using the Equations (2.2.7) and (2.2.8) we show in this section that the meeting point does not exist for this case. These Equations reduce to the following form:

$$\sum_{i=1}^{i=2} \left[W_{ac}(p_i) - W_{rj}(p_i) - c/p_i \right] p_i\, a_i\, (p_i)^{r_m} (1 - p_i)^{n_m - r_m} = 0 \tag{3.1.1}$$

$$\sum_{i=1}^{i=2} \left[W_{ac}(p_i) - W_{rj}(p_i) \right] a_i\, (p_i)^{r_m} (1 - p_i)^{n_m - r_m} = 0 \tag{3.1.2}$$

On substituting

$$k_1 \leftarrow \left(W_{ac}(p_2) - W_{rj}(p_2) - c/p_2\right)\left(W_{ac}(p_1) - W_{rj}(p_1) - c/p_1\right)^{-1},$$

$$k_2 \leftarrow \left(W_{ac}(p_2) - W_{rj}(p_2)\right)\left(W_{ac}(p_1) - W_{rj}(p_1)\right)^{-1},$$

$q_1 \leftarrow (1 - p_1)$ and

$q_2 \leftarrow (1 - p_2)$

in (3.1.1) and (3.1.2) and taking logarithms we obtain two simultaneous linear equations in n_m and r_m, as follows.

$$r_m \ln\left[p_1 q_2 / p_2 q_1\right] + n_m \ln\left[q_1 / q_2\right] = \ln\left[a_2 k_1 / a_1\right] \qquad (3.1.3)$$

$$r_m \ln\left[p_1 q_2 / p_2 q_1\right] + n_m \ln\left[q_1 / q_2\right] = \ln\left[a_2 k_2 / a_1\right] \qquad (3.1.4)$$

The above two equations have no solution, because the determinant is 0. Therefore a meeting point cannot be found for the 2-point prior distribution. If the cost of sampling c is dependent on n_m, then k_1 will involve n_m, and the Equations (3.1.3) and (3.1.4) will in general have a solution.

Wetherill [20] points out that a meeting point for the 2-point prior distribution will in general exist if the cost of sampling, c, is a function of n — the number of items sampled. Wetherill finds the meeting point when $c(n) = n^{\gamma}$, and shows that if the meeting point occurs at sample size N, N goes to infinity as $\gamma \rightarrow 0$.

In many practical situations, c will be a constant and therefore a meeting point at a finite N will not be found. In such situations, the decision boundaries may be obtained by the likelihood test ratio procedure given by Vagholkar *et al* [14] will have to be used.

Wetherill [20] also suggests that an approximate solution for the decision boundaries may be found by forcing the meeting point to occur at a "large enough" sample size. We feel that this will require that a heuristic be developed for estimating what the value for this "large" sample size n_m should be. We suggest below one such heuristic for finding a "good" meeting point (n_m, r_m).

Equation (3.1.3) gives a straight line (in the decision space) on which the decision maker is indifferent between acceptance or rejection. Thus to be a meeting point, (n_m, r_m) must lie on this line. Hence for any given n_m there must exist a unique r_m. If $R(p,n_m)$ represents the risk of a sampling plan obtained using (n_m, r_m) as a meeting point when the lot quality is p, then the following (generic) algorithm will produce a close enough estimate for n_m:

heuristic FIND_MEET_PT

/* finds "meeting point" sample size n_m such that the risk of the sampling plan is relatively insensitive to n_m */

BEGIN

$n \leftarrow 0$

repeat $n \leftarrow n + STEP$

/* STEP is a user-specified positive integer */

until $\max\limits_{\substack{p \\ (0 \le p \le 1)}} \left\{ R(p, n) - R(p, n-1) \right\} \le \varepsilon$

/* where ε is a small positive real number */

$n_m \leftarrow n$.

END. /*ofalgorithmFIND_MEET_PT*/

Once a meeting point is forced at (n_m, r_m), numerical procedures as given in Sections (2.3), (2.4), (2.5) and (2.6) can be used to develop the decision boundaries, and hence the sequential sampling plan. We will derive a better technique for forcing the meeting point at a sufficiently large (n_m, r_m) in Section 4.4 (3), when we discuss the sensitivity of the sampling plan to the location of the meeting point.

3.2 EXACT SOLUTION FOR THE 3-POINT PRIOR CASE

The prior distribution in the 3-point prior model is specified as

$$P(p = p_i) = a_i \text{ for } i = 1, 2 \text{ and } 3, \text{ where } \sum_{i=1}^{i=3} a_i = 1.$$

Here p_1, p_2 and p_3 are the three possible values of lot quality.

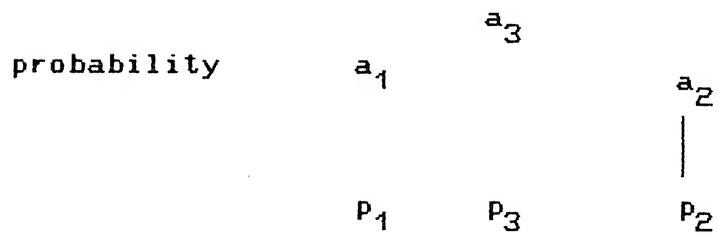

FIGURE 3.2.1. The 3-point prior distribution.

The 3-point prior distribution appears to be an important prior form for the following reasons:

1) It is an improved representation of uncertainty over the 2-point prior of the prior knowledge because it has 5 adjustable parameters, namely the three possible values of the lot quality, and the prior probabilities associated with any two of them (the third is then known).

2) The *analytical* solution for the meeting point can be obtained here as shown in the Section below.

3) The solution of the 3-point prior model enables us to draw some inferences about the behaviour of models based on k-point prior distributions with $k \geq 4$.

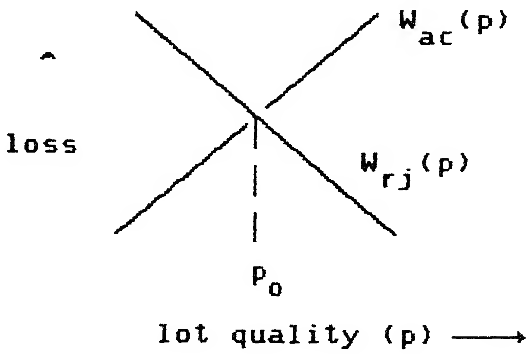

FIGURE 3.2.2. The indifference lot quality, p_o.

We define the indifference quality p_o (Figure 3.2.2) as the lot quality at which the two terminal decisions (accept or reject the lot) will be equally preferred (when the lot quality is $p = p_o$, the risks in accepting or rejecting the lot are equal). If the three p_i's in the three point prior model are all on *one* side of the indifference lot quality p_o, the decision to be made becomes trivial: one would then either always accept or always reject the lot regardless of its quality. Also, if no p_o exists such that $0 \leq p_o \leq 1$, it only means that the losses $W_{ac}(p)$ and $W_{rj}(p)$ have not been correctly quantified for a rational basis for accepting/rejecting the lot to exist.

For the present purpose, we will impose the following restrictions on the selection of p_1, p_2 and p_3:

$$p_1 < p_3 < p_2 \quad \text{and} \quad p_1 < p_o < p_2. \tag{3.2.1}$$

If we additionally locate p_3 at p_o, that provides much computational convenience. In reality, we find that the restriction $p_3 = p_o$ can be relaxed only slightly because as we have found, there is a restriction on the magnitude of the difference between p_3 and p_o: if $|p_3 - p_o|$ exceeds a certain value, the meeting point does not exist.

For notational convenience, we also define the differential loss, $W_d(p)$, as $W_a(p) - W_r(p)$.

3.2.1 THE MEETING POINT

The Equations (2.2.7) and (2.2.8) reduce to the following form for the 3-point prior distribution

$$\sum_{i=1}^{i=3} \left[W_d(p_i) - c/p_i \right] p_i \, a_i \, (p_i)^{r_m} (1 - p_i)^{n_m - r_m} = 0 \tag{3.2.1.1}$$

$$\sum_{i=1}^{i=3} W_d(p_i) \, a_i \, (p_i)^{r_m} (1 - p_i)^{n_m - r_m} = 0 \tag{3.2.1.2}$$

A special case of this situation was solved by Wetherill [19] who showed that the above equations assume a tractable form if one

assumes that the prior has a form such that $p_k(1 - p_k)^{-1} = \lambda^k$, for k = 1, 2 and 3. The equations then become solvable by the Horner's method. However, we find this restriction to be unnecessary. One is able to obtain a closed form analytical solution for the meeting point as follows.

We begin with the following notational substitutions

$$q_i \leftarrow (1 - p_i)$$

$$k_1 \leftarrow (a_2/a_1) \left[W_d(p_2)\right]\left[W_d(p_1)\right]^{-1}$$

$$k_2 \leftarrow (a_3/a_1) \left[W_d(p_3)\right]\left[W_d(p_1)\right]^{-1}$$

$$k_3 \leftarrow (a_2/a_1) \left[W_d(p_2)\ p_2 - c\right]\left[W_d(p_1)\ p_1 - c\right]^{-1}$$

$$k_4 \leftarrow (a_3/a_1) \left[W_d(p_3)\ p_3 - c\right]\left[W_d(p_1)\ p_1 - c\right]^{-1}$$

$$\ell_1 \leftarrow \left[p_2 q_1 \ /\ p_1 q_2\right]^{r_m} \left[p_2 \ /\ q_1\right]^{n_m}$$

$$\ell_2 \leftarrow \left[p_3 q_1 \ /\ p_1 q_3\right]^{r_m} \left[p_3 \ /\ q_1\right]^{n_m}$$

he above substitutions reduce equations (3.2.1.1) and (3.2.1.2) to wo linear equations in X_1 and X_2

$$+ k_1\ X_1 + k_2\ X_2 = 0 \quad \text{and} \qquad (3.2.1.3)$$

$$+ k_3\ X_1 + k_4\ X_2 = 0 \qquad (3.2.1.4)$$

resulting in the solution

$$X_1 = (k_2 - k_4)\,(k_1k_4 - k_2k_3)^{-1} \qquad (3.2.1.5)$$

$$X_2 = (k_3 - k_1)\,(k_1k_4 - k_2k_3)^{-1} \qquad (3.2.1.6)$$

Recall that $\{k_i\}$ are functions of the loss functions, the sampling cost and the prior distribution.)

Back-substituting for X_1 and X_2, and replacing the above solutions (equations 3.2.1.5/6) by k_a and k_b respectively,

$$\left(p_2q_1 \,/\, p_1q_2\right)^{r_m} \left(p_2 \,/\, q_1\right)^{n_m} = k_a \qquad (3.2.1.7)$$

$$\left(p_3q_1 \,/\, p_1q_3\right)^{r_m} \left(p_3 \,/\, q_1\right)^{n_m} = k_b \qquad (3.2.1.8)$$

Taking logarithms on both sides of equations (3.2.1.7) and (3.2.1.8), we obtain the following solution for the meeting point

$$n_m = \left[\ln k_a \ln\left(p_3q_1/p_1q_3\right) - \ln k_b \ln\left(p_2q_1/p_1q_2\right)\right]\left[\Delta\right] \qquad (3.2.1.9)$$

$$r_m = \left[\ln k_b \ln\left(q_2/q_1\right) - \ln k_a \ln\left(p_3/p_1\right)\right]\left[\Delta\right]^{-1} \qquad (3.2.1.10)$$

where,

$$\Delta = \ln\left(p_3q_1/p_1q_3\right)\ln\left(q_2/q_1\right) - \ln\left(p_2q_1/p_1q_2\right)\ln\left(p_3/p_1\right) \qquad (3.2.1.11)$$

3.2.2 CONDITIONS FOR A MEETING POINT TO EXIST

For a meeting point (n_m, r_m) to exist, equations (3.2.1.9) and (3.2.1.10) must have a feasible solution. In addition, n_m and r_m must be greater than 0. The necessary conditions for a meeting point to exist will now be discussed.

Condition 1: $\Delta > 0$

It is easily seen from (3.2.1.11) that Δ is zero if and only if $p_1 = p_2 = p_3$. Thus a condition $\Delta = 0$ in a real decision problem will never occur.

Condition 2: $k_a, k_b > 0$

This condition is required to be satisfied because both k_a and k_b are arguments to logarithms in the equations for n_m and r_m. From equations (3.2.1.5) and (3.2.1.6), it can be seen that $(k_2 - k_4)$, $(k_3 - k_1)$ and $(k_1k_4 - k_3k_2)$ all must have the same sign. The resulting conditions in terms of the losses, the cost of sampling and the prior distribution are easy to derive, and we will here give only an illustration of what these conditions imply. When $p_3 = p_o$, the only constraint to be satisfied for Condition 2 to hold is that

$$(p_2 - p_1) \geq c \left[1/W_d(p_1) + 1/W_d(p_2)\right] \qquad (3.2.2.1)$$

The above inequality places a constraint on the "resolution" of the sampling plan: for a given differential loss function $W_d(p)$ and the cost of sampling c, the exact meeting point cannot be obtained if the difference between p_2 and p_1 is less than what is specified in the above equation. The larger the differential loss $W_d(p)$ and the smaller the cost of sampling, the "finer" is the resolution of the sequential plan in terms of the possible values of p_1, p_2 and p_3.

Condition 3: $n_m, r_m > 0$

We start with making a proposition $\mathbb{P}$ about Δ which we will prove later in this section:

Proposition $\mathbb{P}$: Δ is positive for $p_1 < p_3 < p_2$.

Because we consider our problem only for the case where $p_1 < p_3 < p_2$ (see Equation (3.2.1)), we are assured that the denominator in the equations for n_m and r_m is positive.

Analytically finding the condition for the numerators to also be positive turns out to be a mathematically very complicated problem. We will discuss some of the numerical results obtained in this regard. Taking the cue from the results of Condition 2, for $p_3 = p_0$, if we try to find a relationship between $(p_2 - p_1)$ and $c\left[1/W_d(p_1) + 1/W_d(p_2)\right]$, we find that

$$(p_2 - p_1) > 1.5\ c \left[1/W_d(p_1) + 1/W_d(p_2)\right] \qquad (3.2.2.2)$$

guarantees that $n_m > 0$ and $r_m > 0$, if both $W_d(p_1)$ and $W_d(p_2)$ are greater than $(p_2 - p_1)/(1.5\ c)$. For the special case where $W_d(p_1) = W_d(p_2) = (p_2 - p_1)/(1.5\ c)$, both n_m and r_m are exactly equal to zero.

Proof of Proposition $\mathbb{P}$:

From equation (3.2.1.11), we see that $\Delta = 0$ at $p_1 = p_3$ and $p_2 = p_3$. It follows that $\Delta(p_3)$ has at least one extremum for $p_1 < p_3 < p_2$. On partially differentiating Δ w.r.t p_3, we also find that Δ has only one extremum for $0 < p_3 < 1$. Thus Δ has only *one* extremum for $p_1 < p_3 < p_2$. Now, if Λ_1 and Λ_2 denote the slope of Δ at $p_3 = p_1$ and $p_3 = p_1$ respectively, then

$$\Lambda_1 = \left[\ln(1 - p_2) - \ln(1 - p_1)\right] p_2 - \left[\ln p_1 - \ln p_2\right](1 - p_2) \qquad (3.2.2.3)$$

$$\Lambda_2 = \left[\ln(1 - p_2) - \ln(1 - p_1)\right] p_1 - \left[\ln p_1 - \ln p_2\right](1 - p_1) \qquad (3.2.2.4)$$

Because there is only one extremum for $\Delta(p_3)$ for $p_1 < p_3 < p_2$, $\Lambda_1 > 0$ implies $\Lambda_2 < 0$ and vice versa. From Equations (3.2.2.3) and (3.2.2.4), we get $(\Lambda_1 - \Lambda_2) > 0$, as $p_2 > p_1$. This implies that the slope of Δ at $p_3 = p_1$ is positive and hence Δ itself is positive for $p_1 < p_3 < p_2$.

CHAPTER IV

SENSITIVITY ANALYSIS

4.0 INTRODUCTION

Optimal (sequential) sampling plans (OSPs) obtained from the k-point prior distribution minimize the expected (Bayes) risk of sequential decisions in acceptance sampling as guided by the sequential sampling plans obtained using the theory given in Chapters II and III. In this chapter we will study the effect of changing the various cost, loss and prior distribution parameters on these sampling plans. In addition, we would also study the effect of *imperfect* knowledge of these parameters on the optimality of the sampling plan, as shown in Figure 1.2.2. We will be here be concerned only with the *3-point* prior case, for which the meeting point can be obtained analytically. It is our belief that general k-point prior distributions will give similar results.

The following cases will be studied in this chapter:

(1) The effect of the cost of inspection, on

(a) The expected risk of using the OSP.

(b) The ASN curve for the OSP.

(c) The OC curve for the OSP.

(2) The effect of the choice of the prior distribution, on

(a) The expected risk of using the OSP.

(b) The OC curve of the OSP.

(3) The effect of using a meeting point different from the exact meeting point, on

(a) The decision boundaries of the OSP.

(b) The expected risk of using the OSP.

(4) The effect of imperfect knowledge of the inspection cost, on the expected risk of using the resulting sub-optimal plan.

(5) The effect of imperfect knowledge of the decision losses, on the expected risk of using the resulting sub-optimal plan.

(6) The effect of imperfect inspection, on the expected risk of using the resulting sub-optimal plan.

4.1 THE MODEL

While, in general, any arbitrary cost model can be used for generating OSPs for the k-point prior case, we will here assume a simple case where rejected items are recycled at a fixed unit cost. The monetary loss of accepting a lot is only due to the defective items that remain in it after inspection, the loss of accepting each

such (defective) item being the expected loss due to "external failure" (if this item is to be used in-house), assumed here to be a constant. The cost of inspecting one item is modeled as a constant.

The decision losses and the inspection cost thus become:

$W_{ac}(p)$ = the expected loss of accepting the lot = $L\ p\ C_{EF}$,
$W_{rj}(p)$ = the expected loss of rejecting the lot = $L\ C_{RECYC}$, and,
c = the cost of sampling = C_s,

where

C_{EF} = the cost of external failure of one item,
C_{RECYC} = the cost of recycling one item,
C_s = the cost of inspecting one item, and,
L = lot size.

At the *indifference* quality level, p_o, the losses resulting from decisions "accept lot" and "reject lot" are equal and the decision maker is indifferent between choosing either of these two decisions. In the above model, the indifference quality is

$$p_o = \left[C_{RECYC} / C_{EF}\right] \tag{4.1.1}$$

The "ideal" loss function is thus (from Equation 2.6.1):

4.3 THE IDEAL LOSS FUNCTION

From Equation 2.6.1, the ideal loss function for the above problem is:

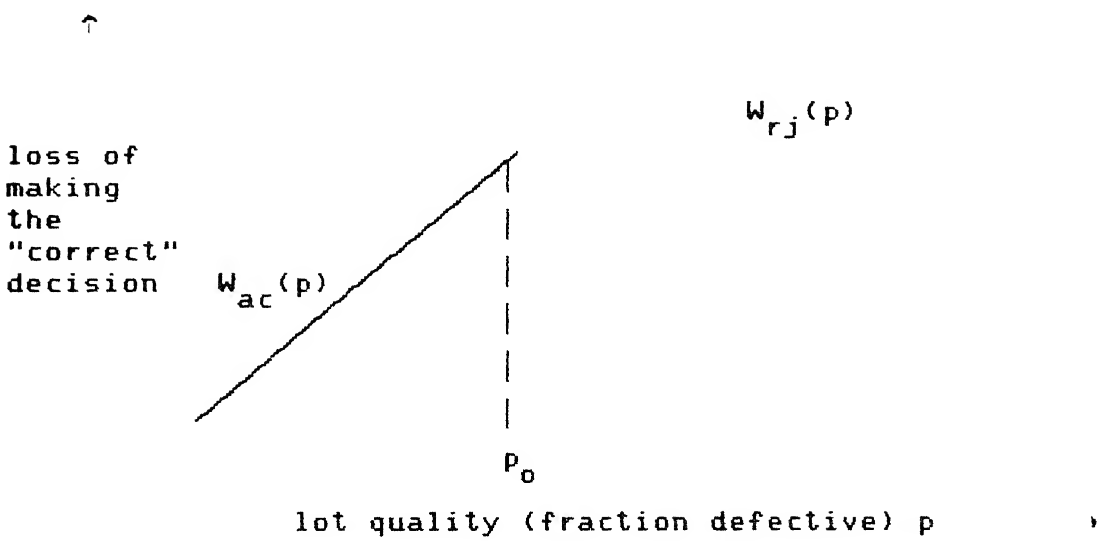

FIGURE 4.1.1. The "ideal" loss function.

An "ideal" loss function can only be realized when the lot quality can be known perfectly, without incurring *any* inspection cost.

4.3 NUMERICAL RESULTS FOR AN OPTIMAL SSP.

(Tabulated in TABLE A.1 in Appendix A.)

In this section we present the numerical results obtained for an optimal SSP for the following example:

Losses and inspection cost:

C_{EF} = 100

C_{RECYC} = 10

C_s = 1

thus, the indifference quality level $p_o = 10/100 = 0.1$

Prior distribution (FIGURE 4.3.1):

p_1 = 0.04

p_2 = 0.2

p_3 = 0.1 (fixed at p_o, the indifference quality level)

$a_1 = a_2 = a_3 = 1/3$ (i.e., p_1, p_2 and p_3 are equally likely)

probability	1/3 a_1	1/3 a_3	1/3 a_2
	0.04 p_1	0.1 p_3	0.2 p_2

FIGURE 4.3.1. The 3-point prior distribution.

Using the analytical results for the meeting point in Section 3.2.1, the meeting point of the acceptance and the rejection boundaries for this example is obtained at (202.590, 20.330). The effective sample size, N^* (Section 2.3.1c), is subsequently obtained at a sample size of 176, where (176, 17) is an acceptance point and (176, 18) a rejection point. The decision boundaries obtained using the algorithm BOUNDARIES are shown in Figure 4.3.2.

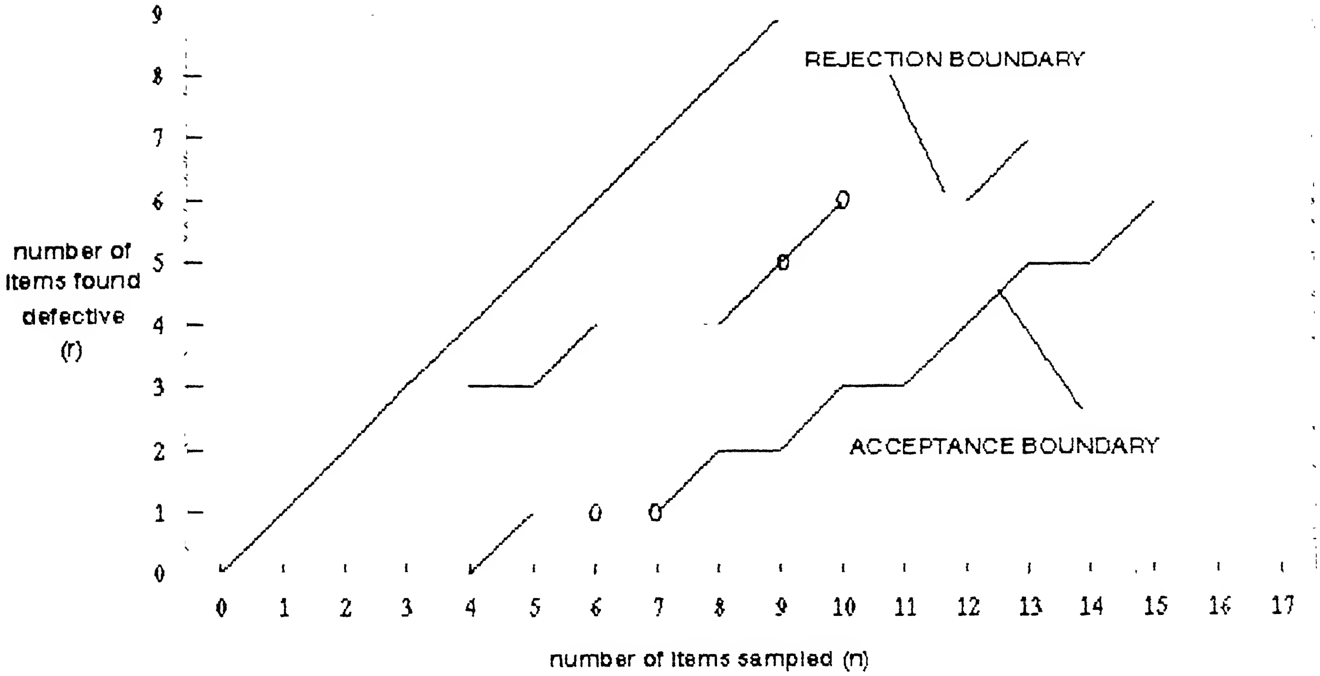

FIGURE 4.3.2.

sample size of 176, where (176, 17) is an acceptance point and (176, 18) a rejection point. The decision boundaries obtained using the algorithm BOUNDARIES are shown in Figure 4.3.2.

Once the above boundaries have been obtained, the calculation of the ASN and OC curves can proceed as stated in Section 2.5.2, using equations 2.5.2.1 and 2.5.2.2. The ASN and the OC curves are shown in Figures 4.3.3 and 4.3.4 respectively.

The expected loss of using the OSP shown in Figure 4.3.2 can now be obtained using the results of Section 2.6. The expected loss

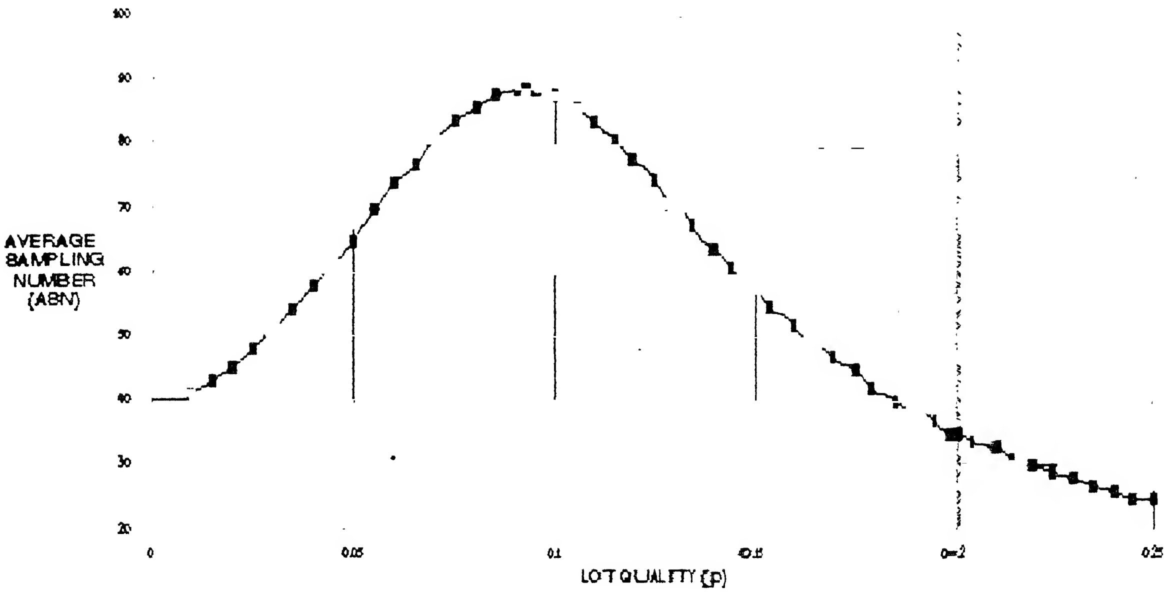

FIGURE 4.3.3.

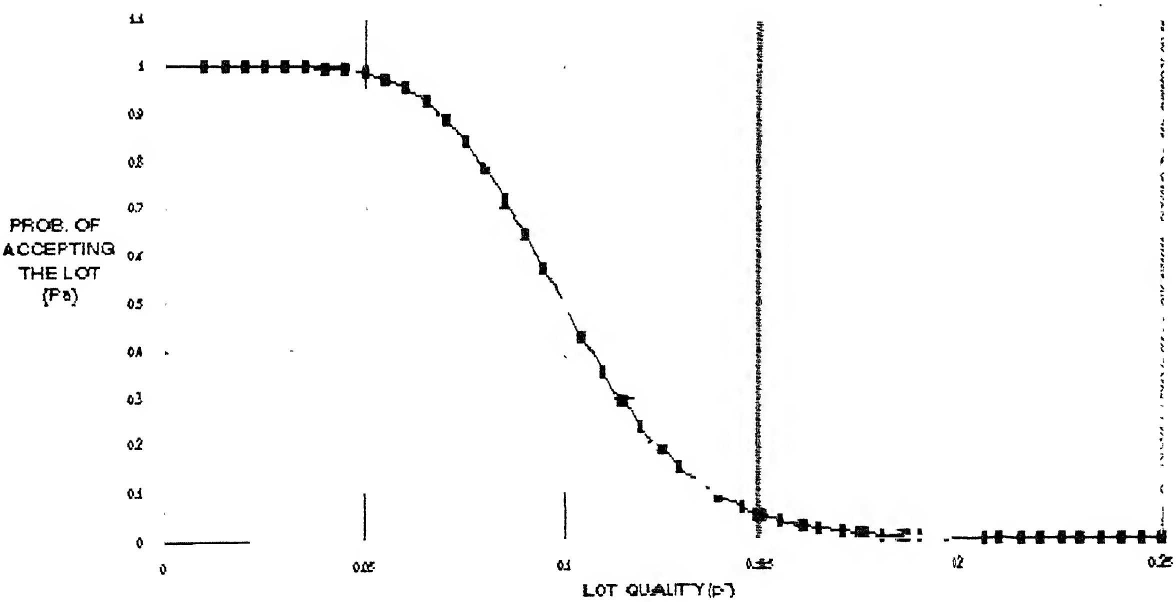

FIGURE 4.3.4.

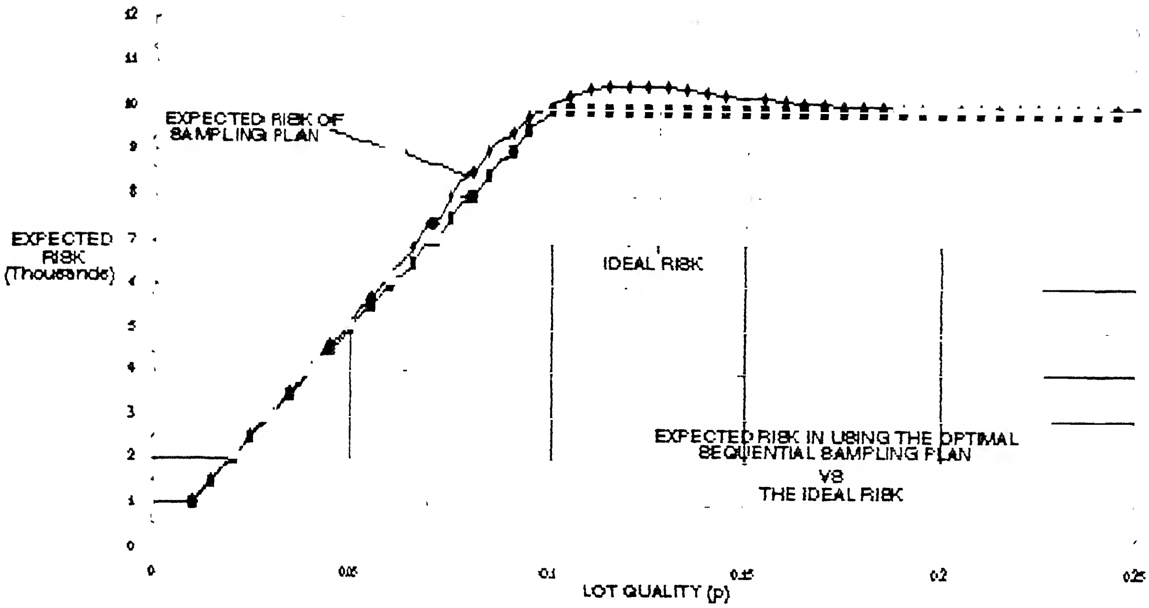

FIGURE 4.3.5.

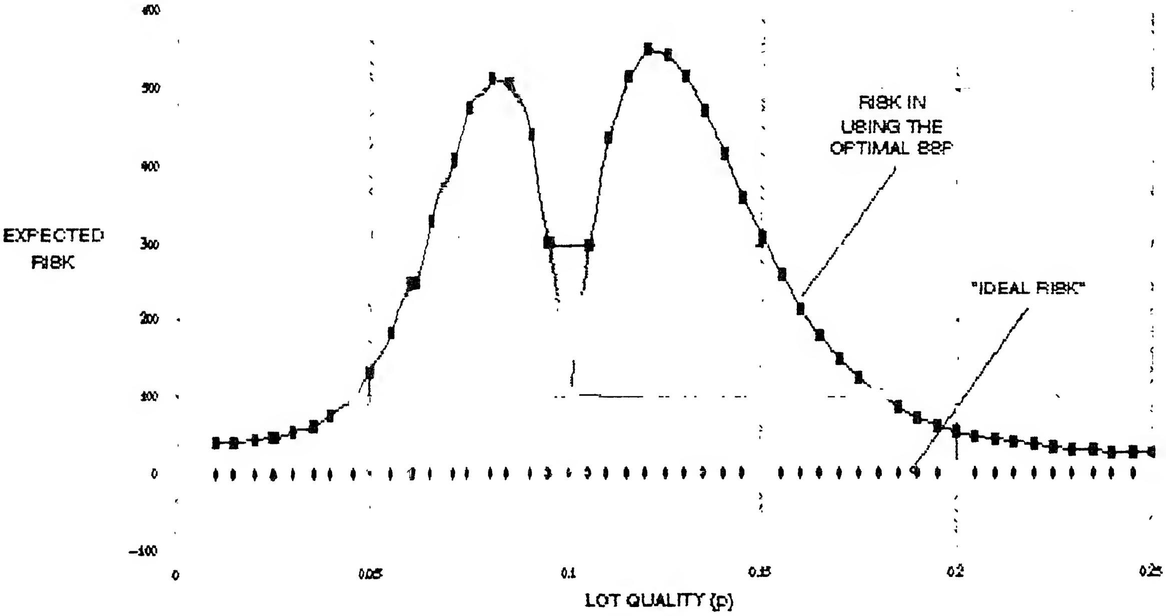

FIGURE 4.3.6.

of using the OSP is compared with the "ideal" loss in Figure 4.3.5. The loss due to imperfect information, δ, which is also the loss of wrong decision, is shown in Figure 4.3.6.

4.4 SENSITIVITY ANALYSIS

(1) Effect of the cost of inspection.

(1a) Expected Risk. (Figure 4.4.1.)

The expected risk goes down as the cost of inspection decreases. Reduced cost of inspection permits the decision maker to obtain a larger sample size, and thus improve his knowledge of the lot quality. (TABLE A.2 in Appendix A.)

(1b) The ASN curve. (Figure 4.4.2.)

The Average Sampling Number increases with decreasing cost of inspection because information about the lot quality can be obtained at a lesser cost. (TABLE A.3 in Appendix A.)

(1c) The OC curve. (Figure 4.4.3.)

The OC curve shifts towards the ideal because more information about the lot quality is available at a lesser cost.
(TABLE A.3 in Appendix A.)

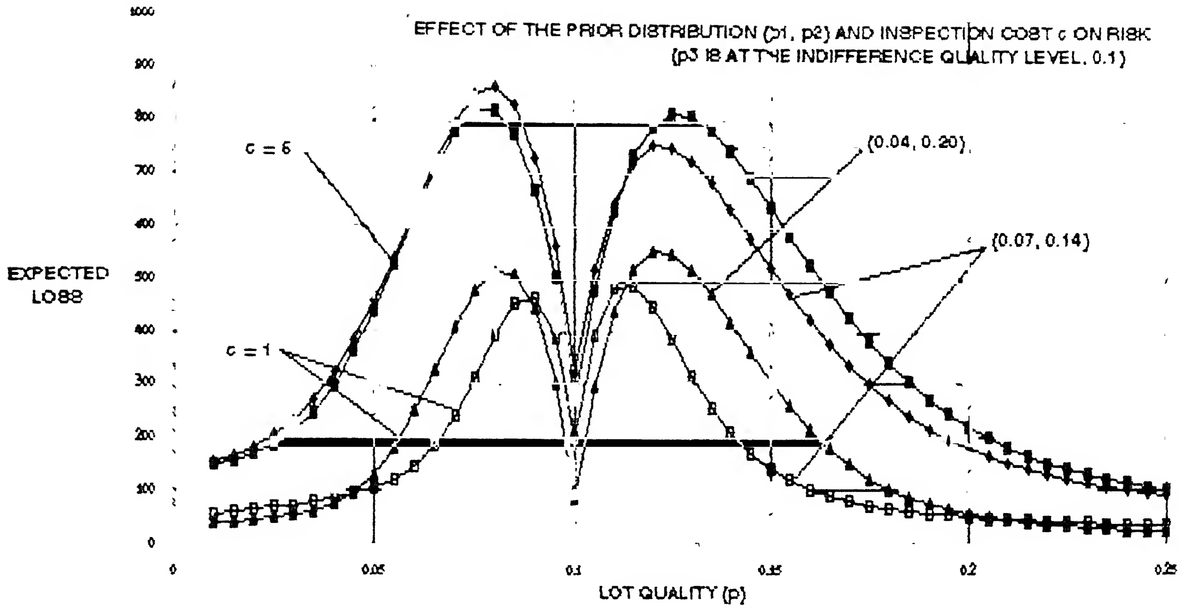

FIGURE 4.4.1

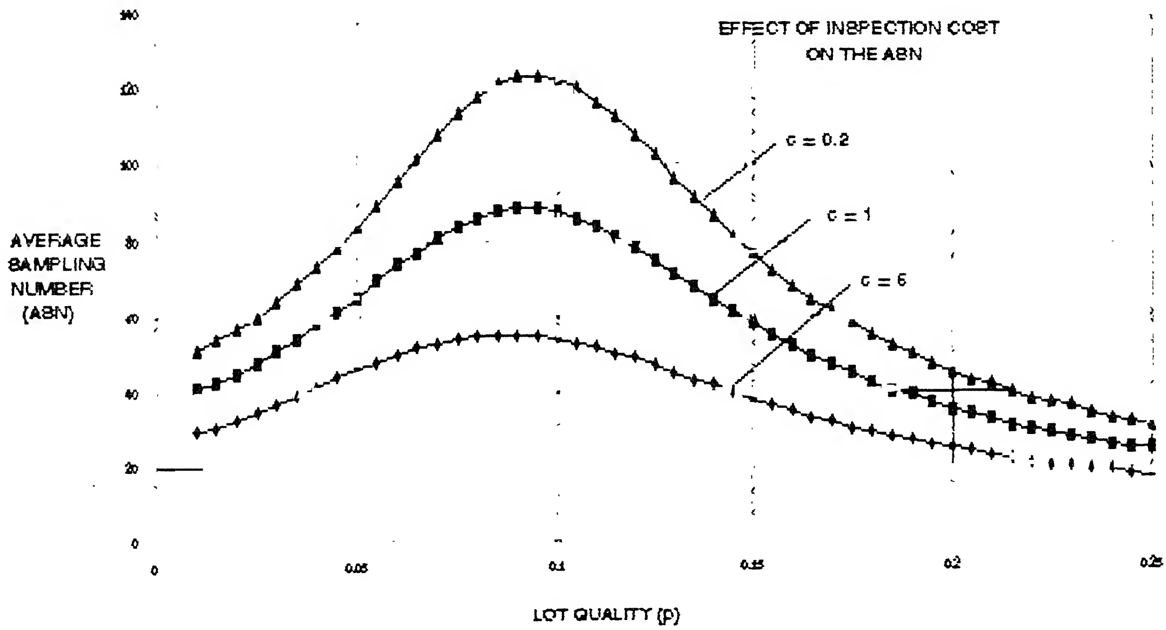

FIGURE 4.4.2

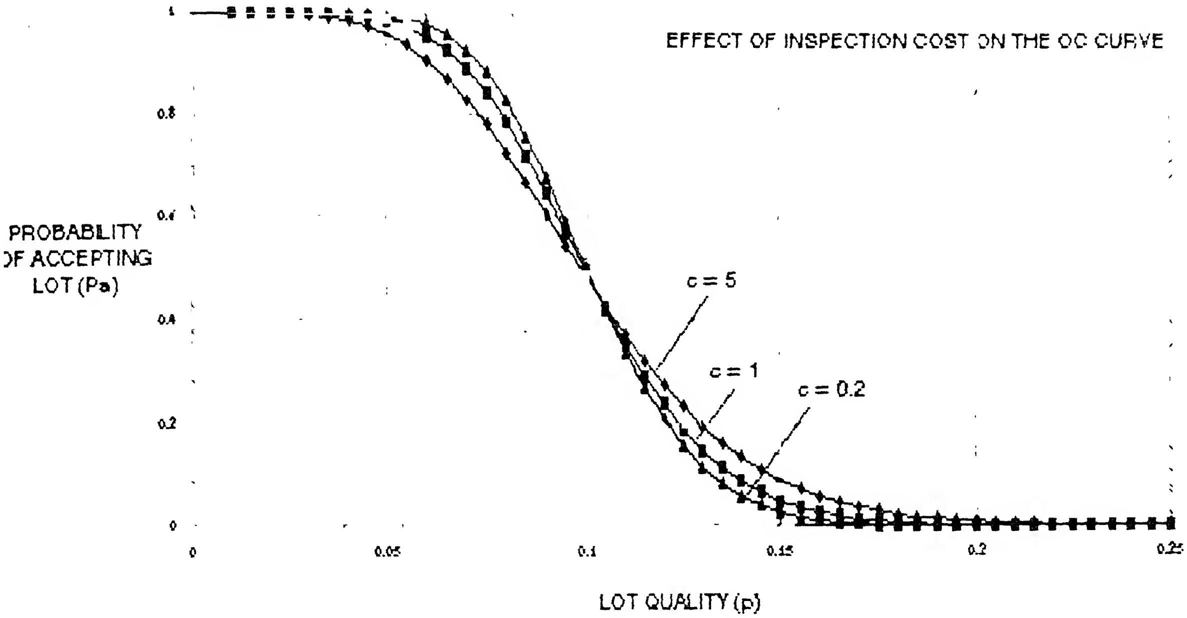

FIGURE 4.4.3.

(2) Effect of the choice of the prior distribution.

(2a) The expected risk. (Figure 4.4.1)

At a lower value of the cost of inspection ($c = 1$), the expected risk is more sensitive to the choice of the prior distribution than at higher values of the cost of inspection. (TABLE A.2 in Appendix A.)

(2b) The OC curve. (Figure 4.4.4)

The OC curve becomes more discriminating as $|p_2 - p_1|$ becomes smaller, as suggested by intuition. But at the same time, the average sampling number increases. (TABLE A.3 in Appendix A.)

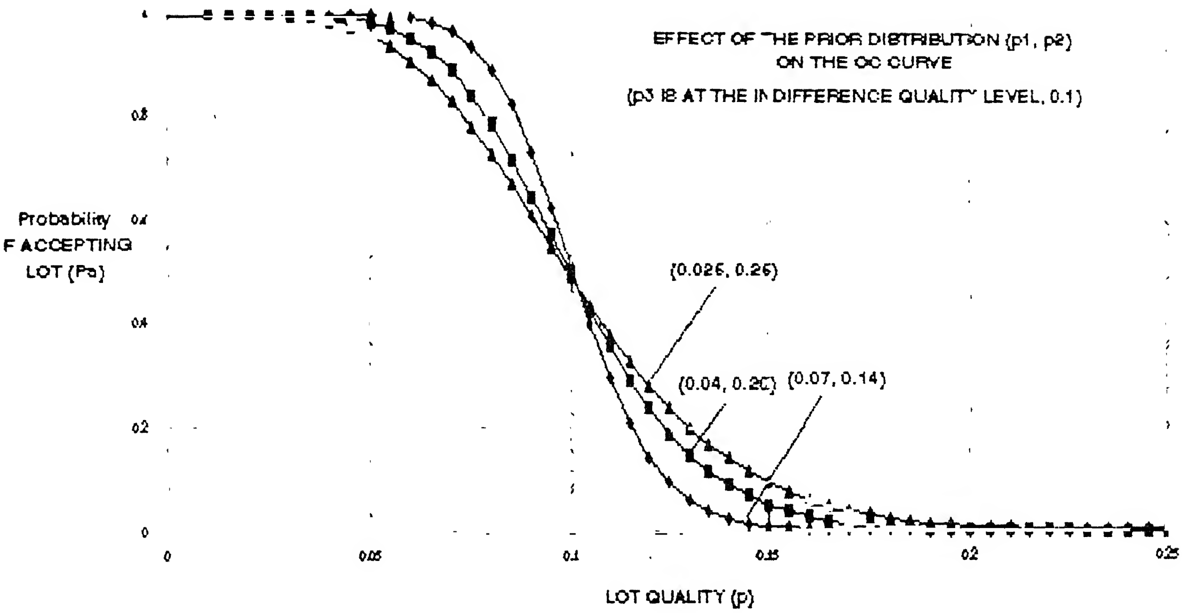

FIGURE 4.4.4.

3) The effect of using a different meeting point

(Figures 4.4.5 and 4.4.6), (TABLE A.6(b) in Appendix A.)

3a) The decision boundaries. (Figure 4.4.5.)

If the exact meeting point cannot be calculated, *optimal* decision boundaries can still be obtained by assuming the meeting point to exist at a sufficiently large sample size. In fact, it is possible to obtain optimal decision boundaries if the meeting point is assumed to exist at a sample size that is

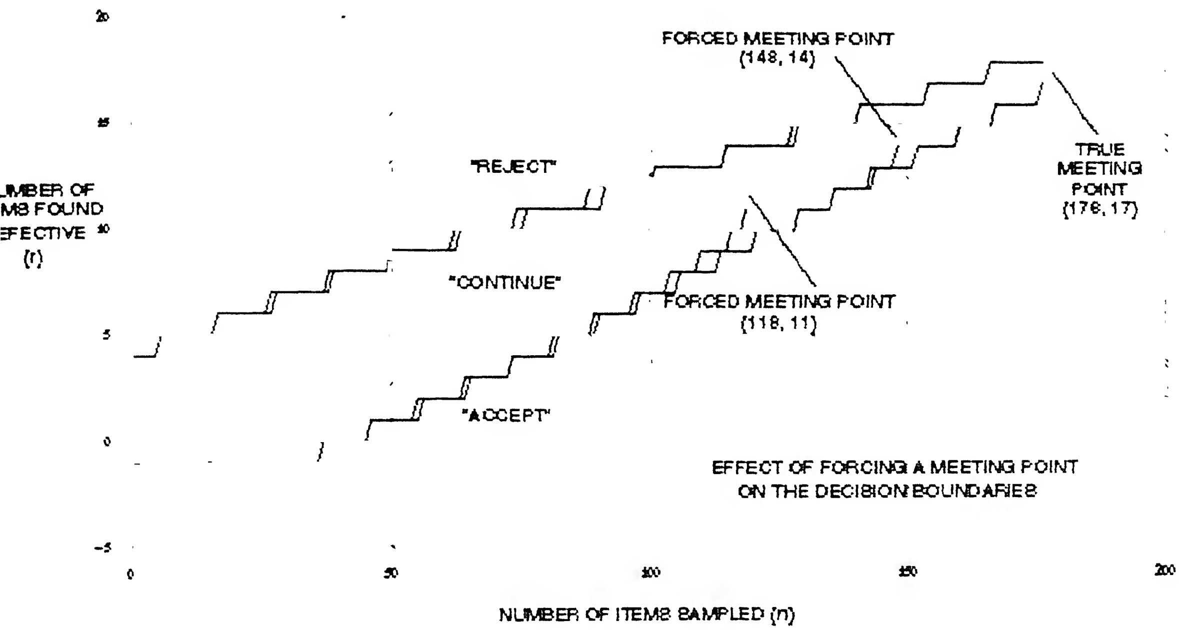

FIGURE 4.4.5.

greater than or equal to the effective sample size, N^*. If the meeting point is forced at (N_f, R_f), then R_f should be chosen as a sufficiently large number, close to N_f to facilitate numerical computations. The above result is a very important one in that it provides for an optimal solution for the OSP even if the meeting point cannot be calculated.

If, on the other hand, the meeting point is forced at a sample size less than N^*, the *total expected* Bayes risk (averaged over the prior distribution) increases, even though the expected risk now is lesser in the vicinity of p_3 (because of a smaller sample size). See Figure 4.4.6. (TABLE A.6(a) in Appendix A.)

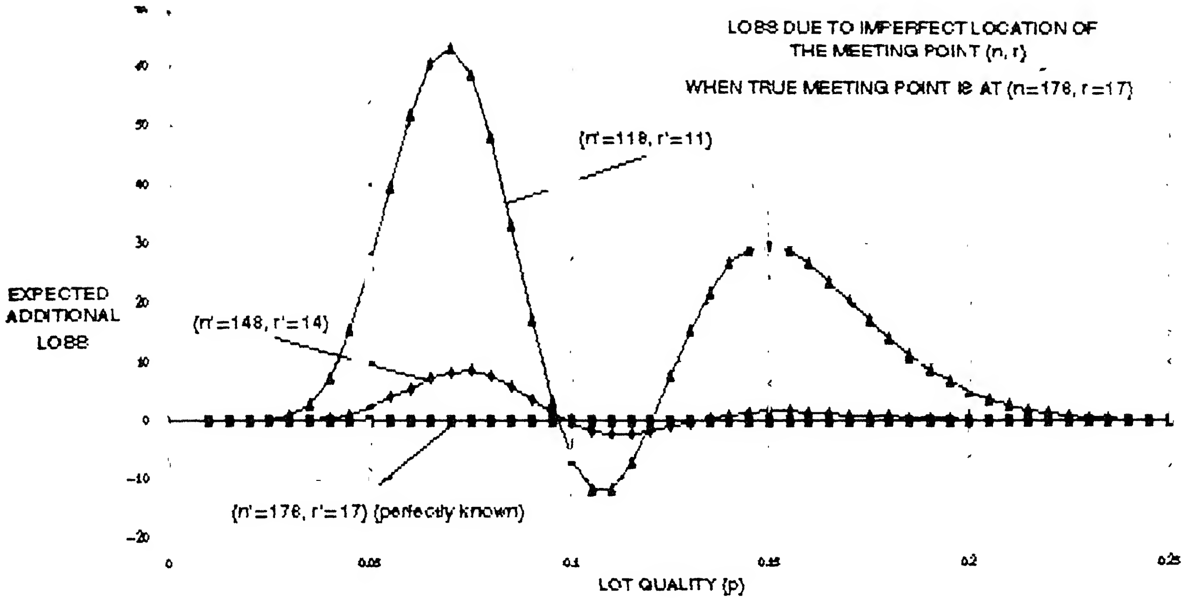

FIGURE 4.4.6

(4) The effect of imperfect knowledge of the inspection cost on the expected risk (Figure 4.4.7.):

As a result of imperfect information about the cost of inspection, the decision boundaries obtained will be different from the optimal decision boundaries. Again, the *total expected* risk of the plan obtained using imperfect information will be higher, even though for a partial set of values of the lot quality, the expected risk will turn out to be lesser.

(TABLE A.4 in Appendix A.)

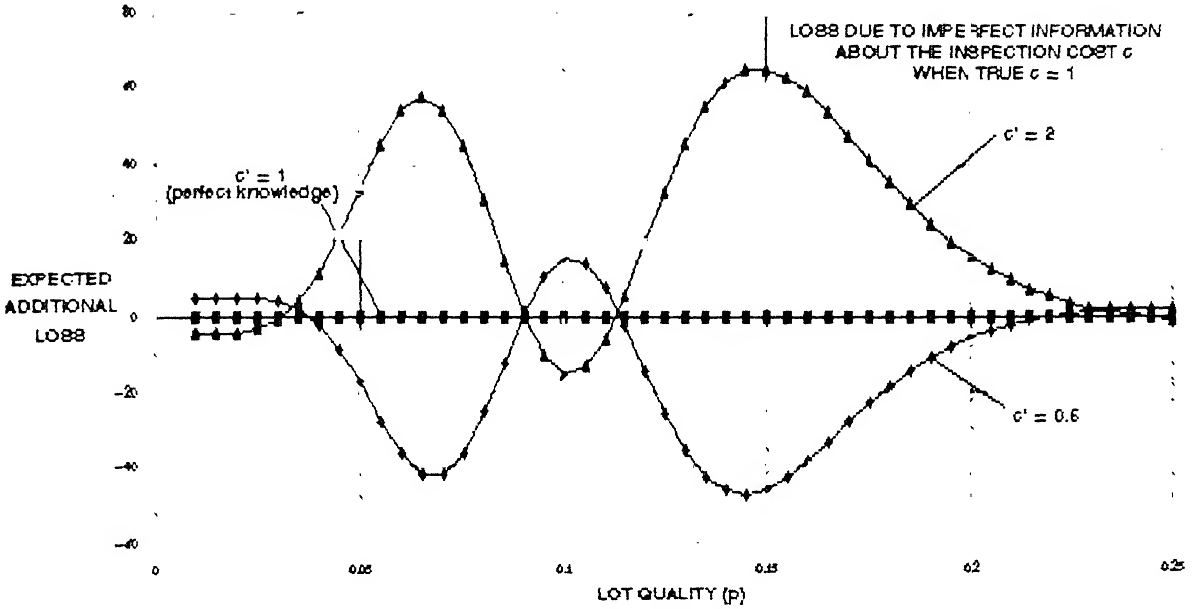

FIGURE 4.4.7.

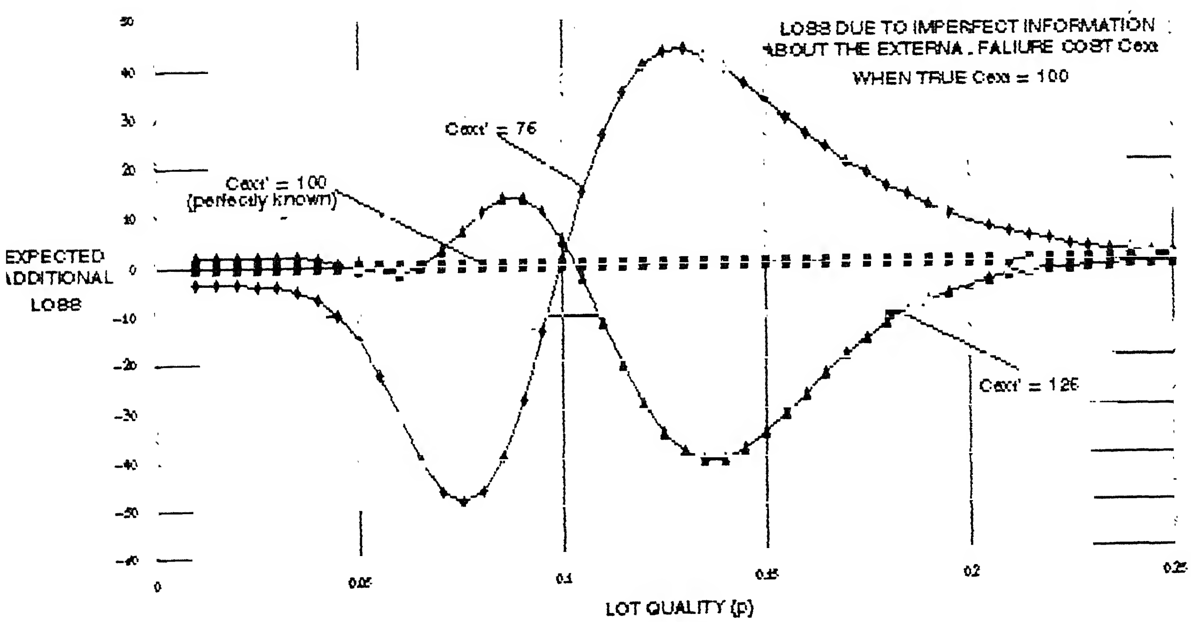

FIGURE 4.4.8.

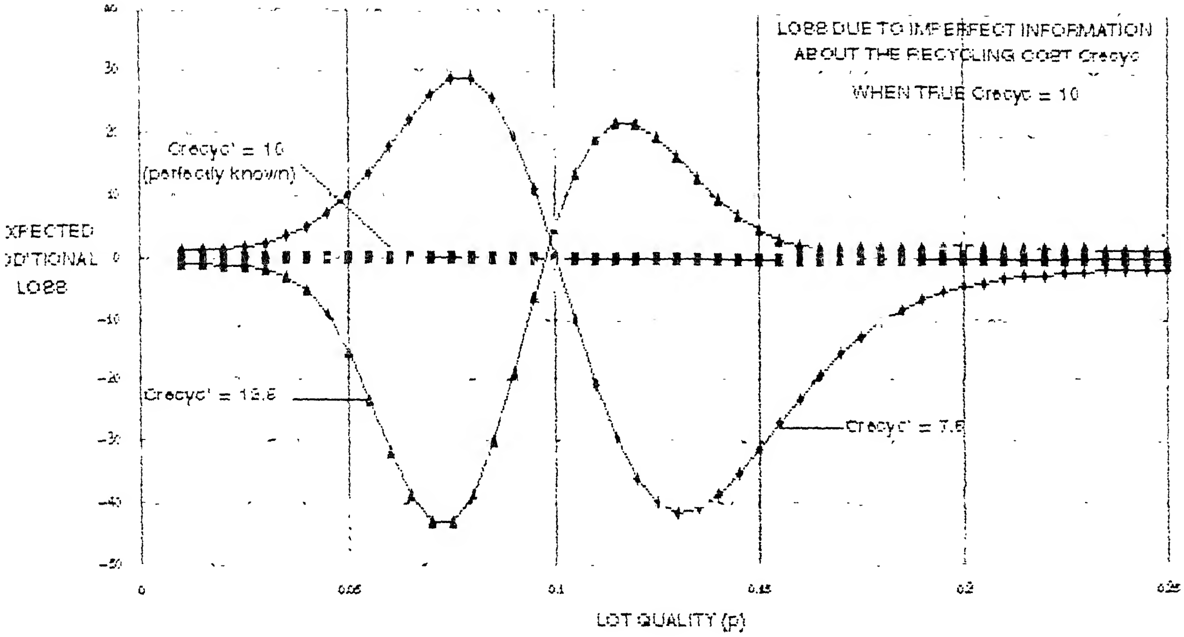

FIGURE 4.4.9

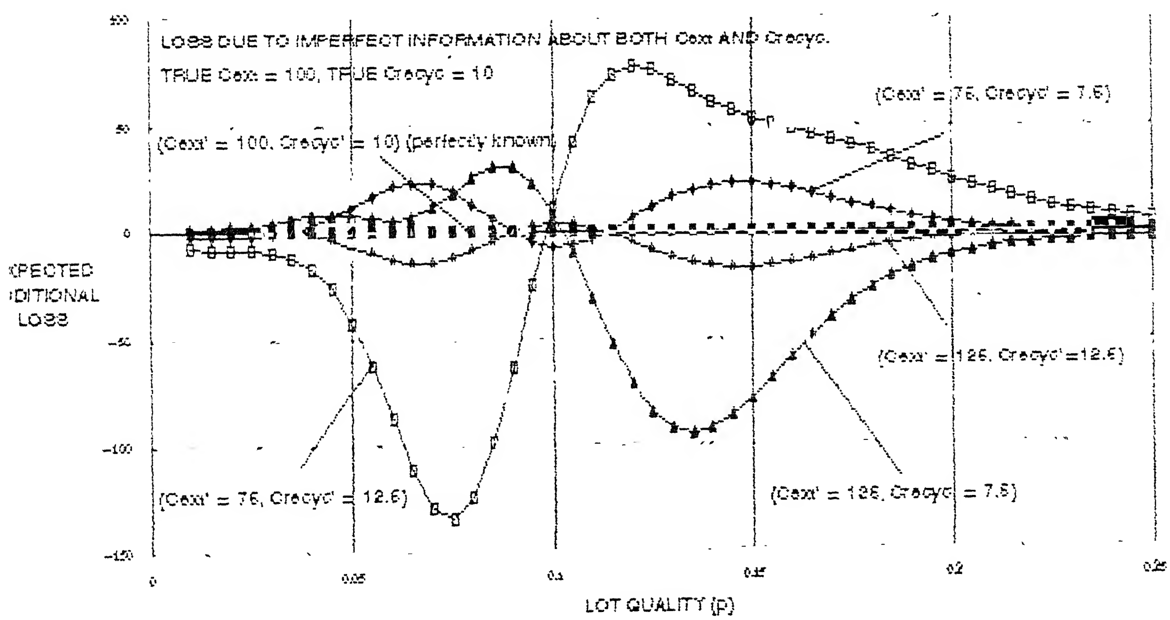

FIGURE 4.4.10

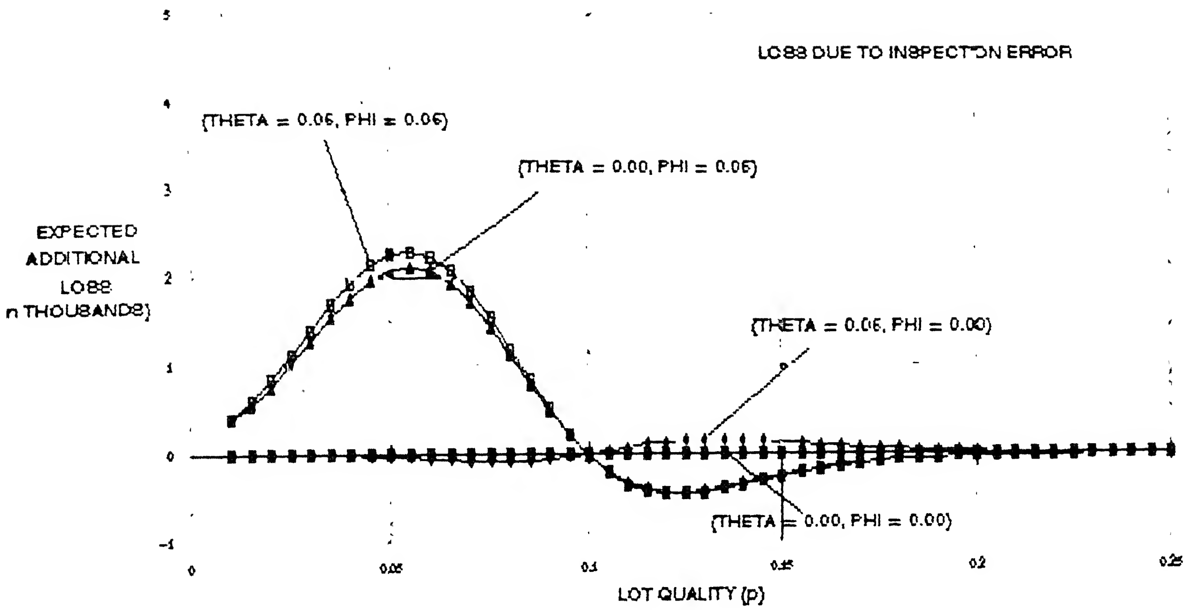

FIGURE 4.4.11.

(5) The effect of imperfect knowledge of the decision losses on the expected risk (Figures 4.4.8, 4.4.9 and 4.4.10.):

(Comments same as in (4).)

(TABLE A.5 in Appendix A.)

(6) The effect of imperfect inspection on the expected risk (Figures 4.4.11 and 4.4.12):

It is observed from the comparison of the effects of Type I and Type II errors, that Type I error (rejection of a good item) results in much larger risks than the Type II error. This is

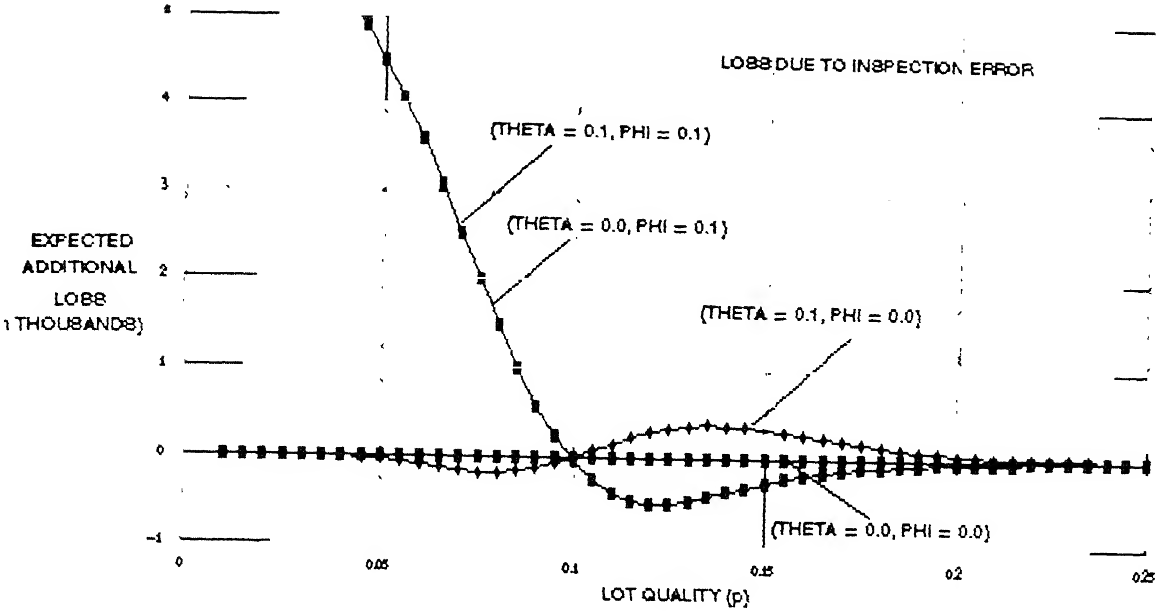

FIGURE 4.4.12

obvious because in a lot of very good quality, even a *very few* incorrectly rejected items will lead to the rejection of the *entire* lot. The obvious implication of this result is the need for accurate inspection equipment and well trained inspectors. In addition, an item should preferably be rejected only after a re-inspection, which will minimize the possibility of a Type I error. Also, in any appraisal of the "quality" of inspection, it should be well understood that the two types of error have vastly different monetary implications.

(TABLE A.7 in Appendix A.)

CHAPTER V

CONCLUSIONS AND DIRECTIONS FOR FUTURE WORK

5.1 CONCLUSIONS

A model was developed to obtain optimal sequential sampling plans (optimized on *expected risk*) based on (1) a generalized discrete prior distribution, called the k-point prior distribution, and (2) the concept of the "meeting point" of decision boundaries separating regions of different decision preferences. Various techniques for locating the meeting point so as to yield optimal plans were demonstrated: (1) numerical solution for $k \geq 4$, (2) analytical solution for $k = 2$ and $k = 3$, and (3) forcing a meeting point at a large enough sample size for all $k \geq 2$. The concept of the "expected risk of a sampling plan" was introduced and the effect of various input factors on the expected risk and the optimality of the plans were studied.

We find that the strength of the k-point prior model lies in the *flexibility* it permits of modeling decision situations which can potentially have any arbitrary loss functions, using the most general (discrete) representation of prior information. Most *practical* problems can be solved satisfactorily using only the 3-point prior distribution which was solved *exactly* to obtain the meeting point.

It was also found (for the example considered) that the optimum obtained is a very "robust" optimum in the sense that the maximum additional loss of under- (or over-) estimating the unit losses and costs by as much as 25% is only about 1% in the *worst case*. On the other hand, a seemingly "small" Type I inspection error of 5% might disturb the optimum by nearly 50% in the worst case. Inspection error of Type II offsets the effect due to the Type I error but only very marginally. Type II inspection errors are in comparison not very important as long as they are not too large.

5.2 APPLICATION OF THE MODEL TO INSPECTION PERFORMANCE.

Inspection error, notably of Type I (rejection of a non-defective item) can very adversely affect the optimality of the sampling plan, as is evident from sensitivity analysis in Chapter IV. We can apply the present (k-point prior) model to optimally track and improve inspection performance, targeting one type of inspection error at a time, as follows.

Suppose we target the Type I error first. For this we set up an inspector audit scheme (i.e., we inspect the inspection process itself) in which the targeted inspector (or inspection station) is fed only good items, "n" representing the cumulative count these good items, without the inspector being aware of it. The number of such items called bad (or defective) by the inspector is the reject count "r" and this is plotted on the (n,r) sample space.

Assuming that the inspector is making Type I errors with a frequency of ϕ for every unit sampled, there are two different actions that could be taken: (1) correction of the error level from $\phi \longrightarrow \phi_1$, where ϕ_1 is the minimum error level possible; the additional cost involved here is the cost of retraining the inspector (or re-calibration of the inspection equipment), and, (2) maintenance of a status quo, i.e., no retraining is done and the error level remains at ϕ. Suppose the loss of maintaining a Type I inspection error level of ϕ over a single day (or any other convenient time unit) of inspection is a function $L(\phi)$. Then, the loss caused by action (1) is

$$W_1(\phi) = C_\phi + L(\phi_1), \tag{5.2.1}$$

a constant, and the loss caused by action (2) is

$$W_2(\phi) = L(\phi), \tag{5.2.2}$$

which will be usually be a monotonically increasing function of ϕ.

Additionally, if the cost of sequentially auditing one instance of inspection is known to be c, then the optimization problem becomes that of determining *whether* to retrain the inspector or not, minimizing the expected loss of the consequences of this decision. If the *a priori* knowledge of the unknown inspection error rate ϕ is expressed as a prior distribution, it is easily seen that the problem reduces to exactly the same form as described in Section 1.2

and can be solved by representing the prior distribution as a k-point prior distribution and using the methodology described in Chapter II. An optimal solution for sequentially tracking the Type II error level can be developed on similar lines.

5.3 DIRECTIONS FOR FUTURE WORK.

Perhaps the most critical aspect of the construction of optimal sequential sampling plans is a reliable quantification of the decision losses and the inspection cost. A method for *objective* assessment of losses and costs accruing to the department or to the business unit (for which sampling inspection is carried out) should be developed, perhaps as an accounting information system, which should be sensitive to any changes within and outside the organization which affects the losses/costs. Better methods for pinpointing the location of the (risk-) indifference quality level are needed.

Analytical solutions to the k-point prior model for $k \geq 4$ will be useful in obtaining a greater insight into the problem, even though it is doubtful that these solutions will result in any real saving in terms of computational effort. This is because the method of forcing the meeting point at a large enough sample size will (we presume) almost always be successful in obtaining optimal decision boundaries, though this has yet to be established.

Group sampling procedures have been completely overlooked in this work, and it is clear that the optimal solution to these procedures will amount to solving a very general scenario in acceptance sampling. Note that group sampling procedures will be better than sequential sampling procedures only if the cost of sampling is a concave function with respect to the number of items sampled in a group (of $\geq$ 1 items).

REFERENCES

[1] BARNARD, G. A., 1946. "Sequential tests in industrial statistics." *JRSS*✝, Suppl., 8, 1-26.

[2] BARNARD, G. A., 1954. "Sampling inspection and statistical decisions." *JRSS*, Series B, 16, 2, 151-174.

[3] BELLMAN, R., 1957. *Dynamic Programming.* Princeton University Press, Princeton, New Jersey.

[4] BURMAN, J. P., 1946. "Sequential sampling formulae for a binomial population." *JRSS*, Suppl., 8, 98-103.

[5] DeGROOT, MORRIS H., 1970. *Optimal Statistical Decisions.* McGraw-Hill Book Company.

[6] DORRIS, A. L., and FOOTE, B. L., 1978. "Inspection errors and statistical quality control: a survey." *AIIE Tr.*, 10, 183-192.

[7] GRANT, E. L., and LEAVENWORTH, R. S., 1980. *Statistical Quality Control*. McGraw-Hill, New York.

[8] HALD, A., 1981. *Statistical Theory of Sampling Inspection by Attributes.* Academic Press, London.

[9] JOHNSON, N. L., 1961. "Sequential analysis: a survey". *JRSS*, A, 124, 372-411.

[10] KELLY, L. G., 1967. *Handbook of Numerical Methods and Applications.* Addision-Wesley Publishing Company.

✝

[11] LECHNER, J. A., 1962. "Optimum decision procedures for a poisson process parameter". *Ann. Math. Stat.*, 33, 1384-1402.

[12] LINDLEY, D. V., and BARNETT, B. N., 1965. "Sequential sampling: Two decision problems with linear losses for binomial and normal variables". *Biometrika*, 52, 507-532.

[13] PRATT, JOHN W., RAIFFA, H., and SCHLAIFER, R., 1965. *Introduction to Statistical Decision Theory.* McGraw-Hill Book Company.

[14] VAGHOLKAR, M. K., and WETHERILL, G. B., 1960. "The most economical binomial sequential probability ratio test". *Biometrika*, 47, 1 and 2, 103-109.

[15] WALD, A., 1943. *Sequential Analysis.* John Wiley, New York.

[16] WALD, A., 1950. *Statistical Decision Functions.* John Wiley, New York.

[17] WALD, A., and WOLFOWITZ, J., 1948. "Optimum character of the sequential probability ratio test". *Ann. Math. Stat.*, 19, 326-339.

[18] WETHERILL, G. B., 1959. "The most economical sequential sampling scheme for inspection by variables". *JRSS*, B, 21, 400-408.

[19] WETHERILL, G. B., 1961. "Bayesian Sequential Analysis". *Biometrika*, 48, 3 and 4, 281-292.

[20] WETHERILL, G. B., 1986. *Sequential Methods in Statistics.* Champan and Hall, London.

APPENDIX A

DATA TABLES FOR SENSITIVITY ANALYSIS

TABLE A.1: RESULTS FOR THE EXAMPLE IN SECTION 4.3

lot quality (p)	ASN	Pa	"ideal" risk R^*	plan risk R	$\delta = R - R^*$
0.01	41	1.0000	1000	1040.74	40.74
0.02	45	1.0000	2000	2045.59	45.59
0.03	51	0.9996	3000	3053.81	53.81
0.04	58	0.9971	4000	4075.16	75.16
0.05	65	0.9865	5000	5133.06	133.06
0.06	74	0.9557	6000	6250.79	250.79
0.07	81	0.8908	7000	7408.64	408.64
0.08	86	0.7855	8000	8515.45	515.45
0.09	89	0.6478	9000	9441.00	441.00
0.10	88	0.4971	10000	10087.75	87.75
0.11	84	0.3551	10000	10438.93	438.93
0.12	78	0.2370	10000	10552.09	552.09
0.13	71	0.1488	10000	10517.59	517.59
0.14	64	0.0885	10000	10418.19	418.19
0.15	58	0.0502	10000	10308.98	308.98
0.16	52	0.0275	10000	10216.82	216.82
0.17	47	0.0146	10000	10149.07	149.07
0.18	42	0.0076	10000	10103.31	103.31
0.19	39	0.0039	10000	10073.92	73.92
0.20	35	0.0020	10000	10055.50	55.50
0.21	33	0.0010	10000	10043.97	43.97
0.22	30	0.0005	10000	10036.59	36.59
0.23	28	0.0003	10000	10031.70	31.70
0.24	26	0.0002	10000	10028.29	28.29
0.25	25	0.0001	10000	10025.78	25.78

TABLE A.2: EFFECT OF PRIOR DISTRIBUTION AND INSPECTION COST ON RISK

lot quality (p)	INSPECTION COST c; PRIORS (p1, p2, p3=0.1			
	c = 1; (p1, p2) (0.04, 0.20)	c = 1; (p1, p2) (0.07, 0.14)	c = 5; (p1, p2) (0.04, 0.20)	c = 5; (p1, p2) (0.07, 0.14)
0.01	40.74	57.63	149.78	156.46
0.02	45.59	64.45	171.56	183.66
0.03	53.81	73.13	213.64	233.61
0.04	75.16	85.21	297.42	320.97
0.05	133.06	105.37	438.07	456.07
0.06	250.79	147.78	618.83	630.24
0.07	408.64	241.18	777.38	796.60
0.08	515.45	390.53	822.06	864.99
0.09	441.00	463.84	668.33	730.57
0.10	87.75	201.53	270.38	324.47
0.11	438.93	482.08	631.17	646.04
0.12	552.09	449.47	791.62	752.93
0.13	517.59	319.82	810.28	725.24
0.14	418.19	209.60	744.87	633.33
0.15	308.98	140.90	641.18	524.11
0.16	216.82	102.05	529.79	422.16
0.17	149.07	79.89	427.65	336.94
0.18	103.31	66.40	341.94	269.74
0.19	73.92	57.48	273.85	218.39
0.20	55.50	51.08	221.58	179.73
0.21	43.97	46.19	182.23	150.73
0.22	36.59	42.28	152.87	128.91
0.23	31.70	39.04	130.94	112.35
0.24	28.29	36.29	114.44	99.61
0.25	25.78	33.91	101.87	89.67

TABLE A.3: EFFECT OF (1) COST OF INSPECTION AND (2) PRIOR DISTRIBUTION ON THE ASN AND THE OC CURVES

lot quality (p)	EFFECT OF COST WITH PRIOR {0.04, 0.10, 0.20} FOR DIFFERENT VALUES OF c						EFFECT OF PRIOR DISTRIBUTION WITH c = 1						lot quality (p)
	ASN			OC			ASN			OC			
	0.2	1.0	5.0	0.2	1.0	5.0	{0.025, 0.25}	{0.04, 0.20}	{0.07, 0.14}	{0.025, 0.25}	{0.04, 0.20}	{0.07, 0.14}	
0.01	51	41	30	1.0000	1.0000	1.0000	33	41	58	1.0000	1.0000	1.0000	0.01
0.02	57	45	33	1.0000	1.0000	0.9993	36	45	64	0.9996	1.0000	1.0000	0.02
0.03	64	51	37	1.0000	0.9996	0.9959	40	51	73	0.9970	0.9996	1.0000	0.03
0.04	73	58	41	0.9994	0.9971	0.9849	44	58	84	0.9865	0.9971	0.9998	0.04
0.05	84	65	46	0.9954	0.9865	0.9582	48	65	99	0.9596	0.9865	0.9988	0.05
0.06	96	74	50	0.9779	0.9557	0.9076	52	74	119	0.9083	0.9557	0.9929	0.06
0.07	108	81	53	0.9278	0.8908	0.8292	54	81	146	0.8293	0.8908	0.9682	0.07
0.08	118	86	55	0.8277	0.7855	0.7261	56	86	175	0.7267	0.7855	0.8920	0.08
0.09	124	89	55	0.6789	0.6478	0.6076	56	89	197	0.6099	0.6478	0.7329	0.09
0.10	123	88	54	0.5053	0.4971	0.4859	55	88	202	0.4902	0.4971	0.5099	0.10
0.11	117	84	52	0.3402	0.3551	0.3721	53	84	187	0.3780	0.3551	0.2946	0.11
0.12	108	78	49	0.2077	0.2370	0.2739	50	78	162	0.2805	0.2370	0.1435	0.12
0.13	97	71	45	0.1159	0.1488	0.1947	47	71	136	0.2007	0.1488	0.0614	0.13
0.14	87	64	42	0.0597	0.0885	0.1342	43	64	113	0.1391	0.0885	0.0242	0.14
0.15	77	58	38	0.0287	0.0502	0.0901	40	58	95	0.0935	0.0502	0.0093	0.15
0.16	68	52	35	0.0130	0.0275	0.0593	37	52	81	0.0613	0.0275	0.0035	0.16
0.17	61	47	32	0.0057	0.0146	0.0384	34	47	70	0.0392	0.0146	0.0014	0.17
0.18	55	42	29	0.0024	0.0076	0.0246	31	42	62	0.0246	0.0076	0.0006	0.18
0.19	50	39	27	0.0010	0.0039	0.0156	29	39	55	0.0152	0.0039	0.0002	0.19
0.20	45	35	25	0.0004	0.0020	0.0098	27	35	50	0.0092	0.0020	0.0001	0.20
0.21	42	33	23	0.0002	0.0010	0.0062	25	33	46	0.0056	0.0010	0.0000	0.21
0.22	38	30	21	0.0001	0.0005	0.0039	23	30	42	0.0033	0.0005	0.0000	0.22
0.23	36	28	20	0.0000	0.0003	0.0025	22	28	39	0.0020	0.0003	0.0000	0.23
0.24	33	26	19	0.0000	0.0002	0.0016	21	26	36	0.0012	0.0002	0.0000	0.24
0.25	31	25	17	0.0000	0.0001	0.0010	19	25	34	0.0007	0.0001	0.0000	0.25

TABLE A.4: EFFECT OF IMPERFECT KNOWLEDGE OF THE INSPECTION COST, c

lot quality (p)	ADDITIONAL RISK WHEN TRUE c = 1		
	c = 1.0	c = 0.5	c = 2.0
0.01	0.00	4.41	−4.44
0.02	0.00	4.76	−4.40
0.03	0.00	3.88	−1.03
0.04	0.00	−2.11	11.05
0.05	0.00	−17.43	33.30
0.06	0.00	−36.35	53.83
0.07	0.00	−41.97	54.20
0.08	0.00	−25.26	30.44
0.09	0.00	0.99	−0.05
0.10	0.00	14.96	−14.93
0.11	0.00	7.22	−5.77
0.12	0.00	−14.32	19.22
0.13	0.00	−35.28	45.26
0.14	0.00	−46.24	61.63
0.15	0.00	−45.98	65.34
0.16	0.00	−38.34	59.15
0.17	0.00	−28.05	47.81
0.18	0.00	−18.40	35.39
0.19	0.00	−10.84	24.35
0.20	0.00	−5.56	15.65
0.21	0.00	−2.16	9.33
0.22	0.00	−0.10	4.99
0.23	0.00	1.09	2.13
0.24	0.00	1.72	0.31
0.25	0.00	2.04	−0.82

TABLE A.5: EFFECT OF IMPERFECT KNOWLEDGE OF LOSSES

ADDITIONAL RISK OF IMPERFECT KNOWLEDGE OF LOSSES WHEN TRUE Cext = 100, Crecyc = 10

lot quality (p)	perfect knowledge of {Cext,Crecyc}:	imperfect knowledge of {Cext}:		imperfect knowledge of {Crecyc}:		imperfect knowledge of both {Cext and Crecyc}:				lot quality (p)
	{100,10}	{75}	{125}	{7.5}	{12.5}	{75,7.5}	{125,7.5}	{75,12.5}	{125,12.5}	
0.01	0.00	−3.30	2.12	1.03	−1.10	−1.95	2.16	−7.42	1.37	0.01
0.02	0.00	−3.64	2.21	1.19	−1.26	−1.89	2.80	−8.04	1.55	0.02
0.03	0.00	−4.18	2.19	2.16	−1.95	−0.77	4.93	−9.68	1.17	0.03
0.04	0.00	−6.62	1.27	4.83	−5.39	3.68	7.84	−17.16	−1.10	0.04
0.05	0.00	−15.19	−0.99	9.99	−15.43	12.61	7.82	−41.76	−6.55	0.05
0.06	0.00	−31.34	−1.72	17.93	−31.40	21.66	6.03	−87.63	−12.90	0.06
0.07	0.00	−46.43	3.01	26.21	−43.00	22.82	11.88	−128.95	−14.46	0.07
0.08	0.00	−46.83	10.95	28.59	−38.54	13.49	26.05	−123.96	−8.65	0.08
0.09	0.00	−28.08	13.54	19.64	−18.54	0.55	30.87	−64.25	0.11	0.09
0.10	0.00	0.73	4.90	0.63	4.70	−6.37	9.95	12.20	4.75	0.10
0.11	0.00	26.13	−12.10	−20.67	19.14	−3.34	−31.31	63.12	2.23	0.11
0.12	0.00	40.33	−29.08	−35.99	21.64	6.30	−71.64	77.93	−4.97	0.12
0.13	0.00	43.41	−39.11	−41.57	16.53	16.36	−93.49	71.68	−12.23	0.13
0.14	0.00	39.56	−40.39	−38.72	9.71	22.46	−93.41	61.36	−16.35	0.14
0.15	0.00	32.97	−35.24	−31.26	4.70	23.49	−78.59	54.00	−16.71	0.15
0.16	0.00	26.20	−27.26	−22.82	2.24	20.75	−58.64	49.02	−14.41	0.16
0.17	0.00	20.24	−19.31	−15.59	1.54	16.22	−40.27	44.09	−10.99	0.17
0.18	0.00	15.32	−12.84	−10.29	1.61	11.50	−26.28	38.17	−7.60	0.18
0.19	0.00	11.36	−8.18	−6.79	1.82	7.48	−16.84	31.51	−4.81	0.19
0.20	0.00	8.24	−5.10	−4.63	1.90	4.44	−10.93	24.92	−2.77	0.20
0.21	0.00	5.84	−3.16	−3.35	1.82	2.33	−7.42	19.04	−1.39	0.21
0.22	0.00	4.06	−1.98	−2.60	1.66	0.95	−5.38	14.18	−0.51	0.22
0.23	0.00	2.76	−1.27	−2.14	1.46	0.08	−4.18	10.39	0.02	0.23

TABLE A.6(a)

items sampled (n)	DECISION BOUNDARY AS ACCEPT/REJECT (A, R) COUNT FOR MEETING POINT {R1, R2}		
	{176,17}	{148,14}	{119,11}
1	–, 4	–, 4	–, 4
2	–, 4	–, 4	–, 4
3	–, 4	–, 4	–, 4
4	–, 4	–, 4	–, 4
5	–, 4	–, 4	–, 4
6	–, 5	–, 5	–, 5
7	–, 5	–, 5	–, 5
8	–, 5	–, 5	–, 5
9	–, 5	–, 5	–, 5
10	–, 5	–, 5	–, 5
11	–, 5	–, 5	–, 5
12	–, 5	–, 5	–, 5
13	–, 5	–, 5	–, 5
14	–, 5	–, 5	–, 5
15	–, 5	–, 5	–, 5
16	–, 5	–, 5	–, 5
17	–, 6	–, 6	–, 6
18	–, 6	–, 6	–, 6
19	–, 6	–, 6	–, 6
20	–, 6	–, 6	–, 6
21	–, 6	–, 6	–, 6
22	–, 6	–, 6	–, 6
23	–, 6	–, 6	–, 6
24	–, 6	–, 6	–, 6
25	–, 6	–, 6	–, 6
26	–, 6	–, 6	–, 6
27	–, 7	–, 7	–, 7
28	–, 7	–, 7	–, 7
29	–, 7	–, 7	–, 7
30	–, 7	–, 7	–, 7
31	–, 7	–, 7	–, 7
32	–, 7	–, 7	–, 7
33	–, 7	–, 7	–, 7
34	–, 7	–, 7	–, 7
35	–, 7	–, 7	–, 7
36	–, 7	–, 7	–, 7
37	0, 7	0, 7	0, 7
38	0, 8	0, 8	0, 7
39	0, 8	0, 8	0, 8
40	0, 8	0, 8	0, 8
41	0, 8	0, 8	0, 8
42	0, 8	0, 8	0, 8
43	0, 8	0, 8	0, 8
44	0, 8	0, 8	0, 8
45	0, 8	0, 8	0, 8
46	1, 8	1, 8	1, 8
47	1, 8	1, 8	1, 8
48	1, 8	1, 8	1, 8
49	1, 8	1, 8	1, 8
50	1, 9	1, 9	1, 9
51	1, 9	1, 9	1, 9
52	1, 9	1, 9	1, 9
53	1, 9	1, 9	1, 9
54	1, 9	1, 9	2, 9
55	1, 9	1, 9	2, 9
56	2, 9	2, 9	2, 9
57	2, 9	2, 9	2, 9
58	2, 9	2, 9	2, 9
59	2, 9	2, 9	2, 9
60	2, 9	2, 9	2, 9
61	2, 9	2, 9	2, 9
62	2, 10	2, 10	2, 10
63	2, 10	2, 10	3, 10
64	2, 10	3, 10	3, 10
65	3, 10	3, 10	3, 10
66	3, 10	3, 10	3, 10
67	3, 10	3, 10	3, 10
68	3, 10	3, 10	3, 10
69	3, 10	3, 10	3, 10
70	3, 10	3, 10	3, 10
71	3, 10	3, 10	3, 10
72	3, 10	4, 10	4, 10
73	4, 10	4, 11	4, 10
74	4, 11	4, 11	4, 10
75	4, 11	4, 11	4, 11
76	4, 11	4, 11	4, 11
77	4, 11	4, 11	4, 11
78	4, 11	4, 11	4, 11
79	4, 11	4, 11	4, 11
80	4, 11	4, 11	5, 11
81	4, 11	4, 11	5, 11
82	5, 11	5, 11	5, 11
83	5, 11	5, 11	5, 11
84	5, 11	5, 11	5, 11
85	5, 11	5, 11	5, 11
86	5, 11	5, 11	5, 11
87	5, 11	5, 12	5, 11
88	5, 12	5, 12	5, 11
contd.			
91	6, 12	6, 12	6, 12
92	6, 12	6, 12	6, 12
93	6, 12	6, 12	6, 12
94	6, 12	6, 12	6, 12
95	6, 12	6, 12	6, 12
96	6, 12	6, 12	6, 12
97	6, 12	6, 12	7, 12
98	7, 12	7, 12	7, 12
99	7, 12	7, 12	7, 12
100	7, 12	7, 12	7, 12
101	7, 13	7, 13	7, 12
102	7, 13	7, 13	7, 12
103	7, 13	7, 13	7, 12
104	7, 13	7, 13	8, 12
105	7, 13	7, 13	8, 12
106	8, 13	8, 13	8, 12
107	8, 13	8, 13	8, 12
108	8, 13	8, 13	8, 12
109	8, 13	8, 13	8, 12
110	8, 13	8, 13	8, 12
111	8, 13	8, 13	8, 12
112	8, 13	8, 13	8, 12
113	8, 13	8, 13	8, 12
114	9, 13	9, 13	8, 12
115	9, 14	9, 14	8, 12
116	9, 14	9, 14	10, 12
117	9, 14	9, 14	10, 12
118	9, 14	9, 14	10, 12
119	9, 14	9, 14	11, 12
120	9, 14	9, 14	
121	10, 14	10, 14	
122	10, 14	10, 14	
123	10, 14	10, 14	
124	10, 14	10, 14	
125	10, 14	10, 14	
126	10, 14	10, 14	
127	10, 14	10, 14	
128	10, 15	10, 14	
129	11, 15	11, 15	
130	11, 15	11, 15	
131	11, 15	11, 15	
132	11, 15	11, 15	
133	11, 15	11, 15	
134	11, 15	11, 15	
135	11, 15	11, 15	
136	12, 15	12, 15	
137	12, 15	12, 15	
138	12, 15	12, 15	
139	12, 15	12, 15	
140	12, 15	12, 15	
141	12, 15	12, 15	
142	12, 15	12, 15	
143	12, 15	13, 15	
144	13, 16	13, 15	
145	13, 16	13, 15	
146	13, 16	13, 15	
147	13, 16	13, 15	
148	13, 16	14, 15	
149	13, 16		
150	13, 16		
151	13, 16		
152	14, 16		
153	14, 16		
154	14, 17		
155	14, 17		
156	14, 17		
157	14, 17		
158	14, 17		
159	14, 17		
160	15, 17		
161	15, 17		
162	15, 17		
163	15, 17		
164	15, 17		
165	15, 17		
166	15, 18		
167	16, 18		
168	16, 18		
169	16, 18		
170	16, 18		
171	16, 18		
172	16, 18		
173	16, 18		
174	16, 18		
175	16, 18		
176	17, 18		

TABLE A.6(b): EFFECT OF INEXACT LOCATION OF THE MEETING POINT

lot quality (p)	ADDITIONAL RISK OF LOCATING THE MEETING POINT AT :		
	(176,17)	(148,14)	(119,11)
0.01	0.00	0.00	−0.08
0.02	0.00	0.00	−0.20
0.03	0.00	0.03	0.65
0.04	0.00	0.45	7.41
0.05	0.00	2.24	26.52
0.06	0.00	5.61	51.87
0.07	0.00	8.30	63.16
0.08	0.00	7.56	48.08
0.09	0.00	3.66	17.41
0.10	0.00	−0.48	−6.79
0.11	0.00	−2.40	−11.71
0.12	0.00	−1.91	−0.52
0.13	0.00	−0.35	15.43
0.14	0.00	0.99	26.66
0.15	0.00	1.56	29.90
0.16	0.00	1.49	26.74
0.17	0.00	1.11	20.52
0.18	0.00	0.70	14.00
0.19	0.00	0.39	8.67
0.20	0.00	0.20	4.93
0.21	0.00	0.09	2.58
0.22	0.00	0.04	1.22
0.23	0.00	0.02	0.49
0.24	0.00	0.00	0.13
0.25	0.00	0.00	−0.04

TABLE A.7: EFFECT OF INSPECTION ERROR

ADDITIONAL RISK DUE TO IMPERFECT INSPECTION WITH {TYPE I, TYPE II} OR {THETA, PHI} ERROR PROBABILITIES

lot quality (p)	{0.00,0.00}	{0.05,0.00}	{0.05,0.05}	{0.00,0.05}	{0.10,0.00}	{0.01,0.01}	{0.00,0.10}	lot quality (p)
0.01	0.00	−0.20	391.08	410.92	−0.41	5607.00	5728.39	0.01
0.02	0.00	−0.56	778.07	841.89	−1.10	5788.77	5967.42	0.02
0.03	0.00	−1.72	1283.30	1405.93	−3.21	5634.88	5814.82	0.03
0.04	0.00	−6.67	1779.09	1951.01	−11.87	5187.05	5334.69	0.04
0.05	0.00	−21.91	2092.52	2281.65	−38.68	4489.11	4593.02	0.05
0.06	0.00	−51.38	2083.04	2251.84	−91.75	3582.64	3646.20	0.06
0.07	0.00	−84.22	1723.61	1846.50	−153.69	2540.30	2573.38	0.07
0.08	0.00	−95.42	1127.42	1197.68	−178.98	1491.60	1504.79	0.08
0.09	0.00	−65.55	488.06	514.03	−127.11	592.06	593.88	0.09
0.10	0.00	0.91	−23.35	−26.68	1.02	−45.33	−49.13	0.10
0.11	0.00	78.34	−327.19	−345.36	161.13	−395.13	−401.12	0.11
0.12	0.00	138.49	−436.83	−459.45	296.54	−510.23	−516.63	0.12
0.13	0.00	166.28	−416.19	−437.43	370.52	−478.77	−484.79	0.13
0.14	0.00	162.86	−334.50	−352.00	377.78	−382.50	−387.88	0.14
0.15	0.00	139.16	−241.19	−254.62	335.99	−276.15	−280.86	0.15
0.16	0.00	107.63	−161.54	−171.47	270.21	−186.46	−190.59	0.16
0.17	0.00	77.29	−103.05	−110.32	201.34	−120.82	−124.47	0.17
0.18	0.00	52.61	−63.99	−69.36	141.66	−76.85	−80.12	0.18
0.19	0.00	34.58	−39.46	−43.52	95.69	−49.00	−51.97	0.19
0.20	0.00	22.34	−24.65	−27.81	63.03	−31.92	−34.65	0.20
0.21	0.00	14.45	−15.84	−18.40	41.12	−21.56	−24.10	0.21
0.22	0.00	9.52	−10.62	−12.76	26.99	−15.24	−17.61	0.22
0.23	0.00	6.49	−7.47	−9.32	18.10	−11.28	−13.52	0.23
0.24	0.00	4.63	−5.51	−7.15	12.56	−8.71	−10.84	0.24
0.25	0.00	3.48	−4.24	−5.72	9.11	−6.97	−9.00	0.25

PROGRAM LISTINGS

Program BOUNDARIES

```
: Program to find the (1) Effective Meeting Point, and (2) The Decision */
Boundaries for the case of a 3-point prior.                            */

'ine MAXN    1000
'ine MAXR    1000
lefine maximum (computationally) possible size of the sample space: a   */
rectangular integer grid {(0,0), (0, MAXR), (MAXN,0), (MAXN,MAXR) }     */

:lude <stdio.h>
include standard I/O header file */

:lude <math.h>
include math header file for calculating pow(), abs() */

at   th1, th2, th3, a1, a2, a3, po1, po2, po3;
(th1, th2, th3) are the 3 points on the prior. NOTE: th1 > th3 > th2 */
(a1, a2, a3) are the corresponding prior probabilities */
(po1, po2, po3) hold the posterior values for (a1, a2, a3) */

at   ra1, rr1, ra2, rr2, ra3, rr3, e1, e2, e3;
(ra1, ra2, ra3) are the losses (risks) of accepting lot at (th1, th2,
)*/
(rr1, ..., ...)  .   .    .    .    . rejecting  .   .   .   .   .

(e1, e2, e3) defined in program */

at   ext, recyc, insp, ks;
ext = cost of external failure, recyc = recycling cost,     */
ks = inspection cost. ALL COSTS ARE per item.               */

at   R[MAXN][MAXR];
risk matrix */

     max_r, lotsize, r, n, d1b[MAXN], d2b[MAXN];
max_r is the defect count at the meeting point,  */
r = defect count, n = number of items sampled    */
array d1b holds the defect count at all n for the REJECTION boundary */
  .   d2b   .   .   .    .    .  .  .  .   . ACCEPTANCE   .      */

E    *out;
ascii file -- contains meeting point, decesion boundaries */

ern double acc();
ern double rej();
ble minrisk();
ble posterior();
functions used */
```

```
in()

        FILE *inp, *extf;
        inp = fopen("inp","r");
        out = fopen("outf","w");
        extf = fopen("inpg","w");
        fscanf(inp, "%d %f %f %f\n%f %f\n",&lotsize, &ext, &recyc, &ks, &a1
2);
        insp *= lotsize; cost *= lotsize; ext *= lotsize; recyc *= lotsize;
        /* costs pertaining to the entire lot */

        th3 = (recyc)/(ext);
        /* break-even or risk-indifference quality level */

        printf("indifference quality is %-5.3f\nEnter p2, p1", th3);
        scanf("%f%f",&th2, &th1);

        ra1 = acc(th1); ra2 = acc(th2); ra3 = acc(th3);
        rr1 = rej(th1); rr2 = rej(th2); rr3 = rej(th3);
        printf (" %-7.3f %-7.3f %-7.3f\n%-7.3f %-7.3f %-7.3f\n%-7.3f\n",
                ra1, ra2, ra3, rr1, rr2, rr3, ks);
        e1 = ra1 - rr1;
        e2 = rr2 - ra2;
        e3 = rr3 - ra3;
        a3 = 1-a2-a1;

        printf ("e1 = %-5.2f  e2 = %-5.2f\n", e1 , e2);
        meeting_point();
        /* function finds meeting point */

        boundaries();
        /* function finds decision boundaries */

 /* main */

eeting_point()
* finds the meeting point for the 3-point prior */

        float   k1, k2, k3, k4, kden, ka, kb, aa, ab, ba, bb;
        float   nr_den, real_n, real_r;

        k1 = - (a2/a1)*(e2/e1);
        k2 =  (a3/a1)*(e3/e1);
        k3 = - (a2/a1)*(e2*th2 + ks)/(e1*th1- ks);
        k4 = (a3/a1)*(e3*th3 - ks)/(e1*th1 - ks);

        kden = (k1*k4 - k2*k3);
        ka = (k3 - k1)/kden;
        kb = (k2 - k4)/kden;

        if ((ka <= 0)||(kb <= 0)) {
                printf("Negative log-domain- ka=%-7.4f kb=%-7.4f\n", ka, |
                exit(1);
        }
        /* if ka or kb <= 0, the logarithms cannot be evaluated */

        aa = (th2/th1) * (1-th1)/(1-th2);
        ab = (th3/th1) * (1-th1)/(1-th3);
```

```
      ba = (1-th2)/(1-th1);
      bb = (1-th3)/(1-th1);

      nr_den = (log(ab)*log(ba)-log(aa)*log(bb));
      /* the denominator of the expression for real_n and real_r */

      real_n = (log(kb)*log(ab)-log(aa)*log(ka))/nr_den;
      real_r = (log(ba)*log(ka)-log(bb)*log(kb))/nr_den;
      n = ceil(real_n); r = floor(real_r);
      /* the meeting point is converted into an integer point */

      printf("n=%-7.3f r=%-7.3f\n", real_n, real_r);
/* meeting_point */

nclude "risk.h"

uble posterior(x,y)
 finds posterior at (n,r)=(x,y) given 3-point priors a1, a2, a3 */
 computationally the most expensive procedure */
t       x, y;

      double tot, r1, r2, r3;
      r1 = a1 * pow(th1,(double)y) * pow(1-th1,(double)(x-y));
      r2 = a2 * pow(th2,(double)y) * pow(1-th2,(double)(x-y));
      r3 = a3 * pow(th3,(double)y) * pow(1-th3,(double)(x-y));
      tot = r1+r2+r3;
      po1 = r1/tot;
      po2 = r2/tot;
      po3 = 1-po1-po2;
/* posterior */

uble minrisk(x, y, MIN, TAG)
 finds the decision with the minimum risk at point (x,y) and */
 assigns the minimum risk value to MIN and the decision type to TAG */
t       x, y;
t       *TAG;
oat     *MIN;


      double  tot, r1, r2, r3;
      double  R1, R2, R3;
      r1 = a1 * pow(th1,(double)y) * pow(1-th1,(double)(x-y));
      r2 = a2 * pow(th2,(double)y) * pow(1-th2,(double)(x-y));
      r3 = a3 * pow(th3,(double)y) * pow(1-th3,(double)(x-y));
      tot = r1+r2+r3;
      po1 = r1/tot;
      po2 = r2/tot;
      po3 = 1-po1-po2;
      /* calculate the values for the posterior distribution at (x,y) */
      R1 = po1*rr1 + po2*rr2 + po3*rr3;
      R2 = po1*ra1 + po2*ra2 + po3*ra3;
      R3 = ks + po1*(th1*R[x+1][y+1]+(1-th1)*R[x+1][y])
              + po2*(th2*R[x+1][y+1]+(1-th2)*R[x+1][y])
              + po3*(th3*R[x+1][y+1]+(1-th3)*R[x+1][y]);
      *MIN = (R1<R2)?((R1<=R3)?(R1):(R3)):((R2<=R3)?(R2):(R3));
      *TAG = (R1<R2)?((R1<=R3)?(1):(3)):((R2<=R3)?(2):(3));
      /* tag puts a "tag" on the decision with min risk */
/* minrisk */
```

```
undaries()

        char     ch;
        int      CNT, remain, cents, thous, i, j, k, max_n, max_y, d1, d2, t
        float    min, RSK;

        d1b[n]=r;
        d2b[n]=r+1;
        /* boundaries at the meeting point */

        for (i=0;i<=r-1;++i){
                posterior(n, i);
                R[n][i] = po1*ra1 + po2*ra2 + po3*ra3;
        }
        for (i=r; i<=n;++i){
                posterior(n, i);
                R[n][i] = po1*rr1 + po2*rr2 + po3*rr3;
        }
        /* the risks at meeting point-x (n,xx) are found */

        max_n = n;
        max_r = r+1;
        /* meeting point is initially at n */

        for (j=n-1;j>=0;--j){
                d2 = -1; d1 = 0;
                for (i=0;i<=max_r;++i){
                        minrisk (j, i, &RSK, &tag);
                        R[j][i] = RSK;
                        d2 = (tag == 2) ? (++d2) : (d2);
                        d1 = (tag == 3) ? (++d1) : (d1);
                }
                d2b[j]=d2;
                d1b[j]=d2+d1+1;
                max_r = d1b[j];
                max_n = (d1 == 0) ? (j) : (max_n);
        }
        printf("\n");
        fprintf(out,"%-5d %-5d\n",max_n,d1b[max_n]);
        n = max_n;
        /* update meeting point */
        printf("%-5d %-5d\n",max_n,d2b[max_n]);
        for (i=1;i<=max_n;++i){
           fprintf(out,"%-5d %-5d %-5d %-5d %-10.3f %-10.3f\n",
                i,d2b[i],d1b[i],d1b[i]-d2b[i]-1, R[i][d2b[i]], R[i][d1b[i
        }
        /* order: sample number; boundaries d2, d1; #of C-points */
        fclose(out);
} /* boundaries */
```

2) Program RISK

```
 * ---------------------------------------------------------------- *
 *              ASN calculation -- output of mp used as input        *
 *          OC curve calculation, as a spin off of ASN calculation   *
 *       RISK calculation, from the knowledge of OC  curve  and  ASN *
 *   Data transfer between program mp and this one is through file "out"  *
 * ---------------------------------------------------------------- *

#define MAX      1000
#include <stdio.h>

FILE     *out, *inp, *asnout;
/* "out" is the output of program mp, and is the input for this program */

float    rr[MAX], ra[MAX], ext, recyc, cost, insp;
int      d1b[MAX], d2b[MAX], n, r, lot;
/* d1b[i] is the  rejection boundary at i */
/* d2b[i] is the acceptance boundary at i */
/* (n,r) is the revised meeting point     */

main()
{
        int i, z;
        /* z is there just to absorb un-needed data in the input file "ou
        float p, incr, start, end;

        out=fopen("outf","r");
        inp = fopen("inp", "r");
        asnout = fopen("af", "w");
        fscanf(out,"%d %d\n",&n, &r);
        for (i=1; i<=n; ++i){
                fscanf(out,"%d %d %d %d %f %f\n", &z, &d2b[i],&d1b[i], &z
&ra[i], &rr[i]);
                d1b[i] += 1; d2b[i] +=1;
                /* because the index of an array in C cannot be negative
        }
        fclose(out);
        fscanf(inp, "%d %f %f %f %f", &lot, &ext, &recyc, &cost, &insp);
        ext *= lot; recyc *= lot; cost *= lot; insp *= lot;
        printf("start, end, increment\n");
        scanf("%f%f%f", &start, &end, &incr);
        for(p=start; p<=end; p+=incr) ASN(p);
        fclose(asnout);
}

#include "risk.h"

ASN(p)
/* this function calculates the ASN at quality level p */
/* and as a by product, the OC curve too */
float    p;
{
        float    total, cum_asn = 0.0, prob[MAX], q, tmp;
        float    pr[MAX][MAX], Pa=0.0, R, AC, RJ, MINRSK;
        /* pr[x][y] = probability of reaching point (x,y) during sequent
            sampling */
```

```
/* Pa is the probability of acceptance at quality level p */

int     i, j;

q = 1-p;
for (j=0; j<=10; ++j) pr[0][j]=0.0;
pr[0][1]=1.0;
d1b[0]=d1b[1];
d2b[0]=0;
total = 0.0;
Pa = 0.0;
/* initialization over */

for (i=1; i<=n; ++i){
        prob[i] = 0.0;
        /* initialize */

        pr[i][0] = 0;
        /* "boundary condition" */

        for (j=1; j<=d1b[i]; ++j){
                pr[i][j] = q*pr[i-1][j] + p*pr[i-1][j-1];
        }
        /* iterative equation for REACHABLE boundary points */

        j = d1b[i];
        /* saves typing effort */

        if ((j > i+1) || (j > d1b[i-1])) pr[i][j] = 0.0;
        /* if rejection boundary unreachable */

        for (j=0; j<=d2b[i]-1; ++j) pr[i][j] = 0.0;
        /* for unreachable points */

        j = d2b[i];

        if (j > 0){

                if (d2b[i] <= d2b[i-1]) pr[i][j] = 0.0;
                /* if acceptance boundary unreachable */
                /* otherwise ... */
                Pa += pr[i][j];
                prob[i] = pr[i][j];
                pr[i][j] = 0.0;
        }
        /* j < 0 are unreachable points, stay clear from them! */

        j = d1b[i];
        prob[i] += pr[i][j];
        pr[i][j] = 0.0;
        /* prob[] accumulates probabilities at ALL termination
           points for the current lot quality p */

        total += prob[i];
        cum_asn += i* prob[i];
} /* repeat for all sampling points */
```

```
AC = acc(p);
RJ = rej(p);
MINRSK = (AC <= RJ) ? (AC) : (RJ);
R = Pa * AC + (1 - Pa) * RJ;
printf("p = %-7.3f  total = %-7.4f  ASN = %-7.0f  Pa = %-7.5f
         R = %-10.3f\n", p, total, cum_asn/total, Pa/total, R);
fprintf(asnout, "%-7.3f %-7.0f %-7.5f %-10.3f %-10.3f %-10.3f\n",
    cum_asn, Pa/total, MINRSK, R-MINRSK, R+cum_asn*insp/lot-MINRSK)
```

DISCUSSION

The following points were raised during the thesis examination:

(1) At the theoretical meeting point, one is indifferent between either accepting or rejecting the lot, or taking an additional observation and subsequently making a terminal accept/reject decision. In real life, should one *randomly* select a terminal decision or should one take one more sample?

Comments:

In reality, the decision boundaries will *not* meet because the meeting point (n_m, r_m) will in general be real (See Figure 4.4.5.). So the situation of randomly selecting a terminal decision will not arise. If such a situation arises, one would probably take one more sample and end up conclusively in one of the two terminal decision regions.

(2) Can a generalization be made that inspection errors will *always* contribute heavily to sub-optimality of the sampling plan?

Comments:

No, such a generalization cannot be made. But for a few representative problems we considered, inspection errors did contribute heavily to sub-optimality. Whether this happens or not depends on the loss/cost structure of the problem.

(3) Will a decreasing $|p_1 - p_2|$ always make the OC curve more discriminating?

Comments:

A decrease in $|p_1 - p_2|$ may not always discriminate between p_1 and p_2 better, but will always discriminate between "good" and "bad" lots better.

(4) What are the possible values of the posterior distribution at the meeting point, when the starting prior for the 3-point case is

1/3 1/3 1/3

0.04 0.1 0.2 p $\longrightarrow$

Comments:

If the risk-indifference lot quality is at 0.1, the posterior probability of p = 0.1 will be high as compared to other values of p at the meeting point:

0.04 0.1 0.2

(5) How does the "additional information" curve for lot quality p vary with the number of items sampled?

Comments:

As sampling proceeds, the *marginal* information about p decreases. Thus the above curve will be a monotonically decreasing curve.

IME - 1991 - M - SIN - RIS